COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Factor Models in China

How well do factor models describe equity returns?

An empirical study on factor models in the Chinese equity market

Master Thesis by

Henrik Aalberg 107906
&
Lukas Schubart 107635

Supervisor: Daniel Wekke Probst
External Lecturer, Department of Accounting
Copenhagen Business School

Master of Science in Applied Economics and Finance (cand.merc. EBA)

September 17, 2018

Number of characters/max: 270265/273000

Number of pages/max: 120/120

Abstract

This master thesis concerns itself with factor effects in the Chinese equity market. In order to answer how well factor models explain equity returns in the Chinese market, this thesis tests the CAPM of Sharpe (1964) and Linter (1965), the three-factor model of Fama and French (1993), the four-factor model of Carhart (1997), and the five-factor model of Fama and French (2015) in the Chinese A-share market. In addition to these models, the individual factors that are a part of the four different models are also assessed. The assessment of models and factor effects are performed by means of descriptive methods, as well as univariate regression, multivariate regression, and correlation methods.

Based on the above-described methods, the models are tested in line with the framework developed by Fama and French (1993, 2015). A comparison of standard measures of fit and the more comprehensive *GRS* tests are performed across the models. The resulting findings are that the four- and five-factor models best fit the A-share market. The factors and their overall interplay seem to explain much of the variation in the Chinese equity market. For the multifactor models tested, three out of four sorts performed are unable to reject the null hypothesis of the *GRS* test, and the average R-squared for each model is approximately 0.92.

For the individual factor effects, it can be seen that the market premium is positive but insignificant, the *Size* premium is strong and significant, and there seems to be a lack of a *B/M*, profitability, and investment premia. Furthermore, a significant medium-term reversal effect is present in the market rather than a momentum effect. These findings contrast comparable research on global markets but are broadly in line with research on the Chinese market.

Contents

Table of figures

Section 1 – Introduction

1.1 Introduction

This master thesis concerns itself with factor effects in the Chinese equity market. The concept of factor models in financial literature can be simplified to econometric models that describe asset returns as a result of linear dependency on underlying drivers or factors. This type of equity models has received much academic attention in the last decades, in particular after the influential paper of Fama and French (1993). It has also become more widely applied among practitioners due to the easily implementable methods suggested by and developed from the said paper, as well as the eased implementation technological development has allowed for. The topic is therefore highly relevant both from an academic and from a practitioner's perspective.

First of all, Fama and French (2015) have recently extended their famous three-factor model from Fama and French (1993) with two additional factors: profitability and investment. This model has been tested thoroughly in developed markets, but not in emerging markets. Therefore, it is interesting to investigate the relative explanatory strength between the three-factor model and the five-factor model in an emerging market.

In addition, by far the largest share of the factor research has been conducted on the U.S. equity market due to its high level of data quality tracing back over a long period. That is why there is still a need for out-of-sample tests to provide evidence for the existence of different factor effects. Tests in the U.S. market allow for sound and strong statistical analysis in many ways, but it also introduces an issue known as "the file drawer problem" (Rosenthal, 1979). That is, only the 5% significant results on a topic get published while the 95% insignificant results on the same topic are stored away in a file drawer. In other words, a statistical relationship that is only significant by chance is accepted (type 1 error). Some of the investigated factors, which was found to be significant in initial research would most likely from a recent perspective not be treated as a factor. Research might have pursued factors that seemed evident by chance at the time of their discovery and are not evident any longer. The Chinese equity market very much provides such an out-of-sample test due to its isolated nature. Therefore, an investigation of the effects at play in this market and a rather unprejudiced view on these effects could be a solid check of the validity of different premia.

Furthermore, the theoretical explanations of observed factor premia in different datasets typically argue either rational or behavioral underlying economic drivers. As for the five-factor model of Fama and French (2015), most tests on factor models and factor-based investment have been conducted in developed markets. Developed markets are usually highly correlated with each other, and the additional insights different markets offer to the underlying reasons of factor premia are therefore limited by this. In this light, a market like the Chinese equity market with a high degree of isolation from global markets and a very low correlation with these markets could be useful for shedding more light on the underlying reasons of factor effects.

From a practitioner's perspective, a description of the Chinese equity market is interesting both for specific asset allocation decisions and for benchmarking active returns. The low correlation Chinese equities have shown with global markets, as well as decades of very strong GDP growth, makes it a compelling case for portfolio managers. Furthermore, it is particularly relevant at a point in time with the inclusion of China A shares in MSCI's Emerging Market Index effective from the 31st of May 2018. The inclusion came as a result of several signs from Chinese regulatory authorities of an increasing willingness to ease restrictions and move the market in the direction of less regulation and more global integration. One of the most notable examples of this is the launch of the "Stock Connect Program," which allows all foreign investors to invest in Chinese A-shares on both Shanghai Stock Exchange and Shenzhen Stock Exchange through the Hong Kong Stock Exchange. At the same time, even though restrictions are being eased in China, there is still a considerable amount present. The signals of the increased opening of the Chinese market are likely to increase the amount of capital inflow to the country, and the combination of distinct market peculiarities, foreign interest, and capital inflow makes a description of the market particularly relevant.

The Chinese equity market has not been covered to a large extent in financial literature, and the published research on this market has typically conflicted with earlier research both regarding methods and findings. In addition, several different databases are utilized for data on the Chinese market, both global and local Chinese databases with different levels of accessibility. This makes it hard to compare research and findings across studies. Therefore, there is a need for further and more profound research on the Chinese market, which makes the nature of this thesis at least to some extent exploratory. Due to all of the reasons above, the focus of this thesis is on the Chinese equity market and factor effects in this market. The focus of the thesis will be further specified and elaborated on in the following sections 1.2 describing the problem statement and 1.3 on delimitations.

1.2 Problem statement

To shed light upon as many of the elements of interest briefly mentioned in the introduction, the following overarching problem statement has been worked out:

How well do factor models describe equity returns?

An empirical study on factor models in the Chinese equity market.

This problem statement naturally steers the thesis in the direction of overall equity models, but also in the direction of investigating the individual parts of the chosen factor models. These two different directions are both a part of the widely covering term factor research. Therefore, they are highly related, but it is still two different aspects of factor research that will be covered.

To break down the broad, overall problem statement into specific parts, the following nine sub-questions have been worked out:

1. How is the overall equity market in China structured, and does the structure of the market imply that general methods applied in global markets should be altered?
2. How has equity factor model research developed over the last 30 years and what is the general framework these models operate within?
3. Which market peculiarities do methods applied in research on the Chinese equity market accommodate for and what factors effects have previously been found?
4. Based on Chinese market features and methods applied in previous research, which methods are best suited for the factor analysis in the Chinese market?
5. What data is necessary for the factor analysis in the Chinese market?
6. What factor effects can be found in the Chinese market?
7. How robust are the factor effects found in the Chinese equity market?
8. How can observed factor effects be explained, and how do the observations relate to prevailing explanations?
9. How well do factor models explain equity returns in the Chinese equity market?

1.3 Outline

The thesis is organized in separate parts to answer each of the sub-questions above, which together should answer the overall problem statement as adequately as possible. First, sub-question one will be elaborated on in section 2 on China and the Chinese equity market. However, implications for the further analysis of the outlined structure will not be concluded on in this part. Only the foundation for the later discussion on alterations to standard methods will be outlined, and an overall evaluation based on aspects touched upon in the section on China and the literature review will be made in the literature review, as well as the methodology and data parts. Next, sub-question two and three are covered by the literature review in section 3. The literature review will be extensive both due to the vast amount of existing factor literature and the several particularities that characterize the Chinese market. The latter makes a specific part on factor model research in the Chinese equity market necessary. After that, sub-question four will be answered in section 4 on methodology, and sub-question five will be answered in section 5 on data.

The parts that are dedicated to the five first sub-questions build the foundation necessary to examine factor effects in the subsequent analysis. After this, the main analysis will be performed, where present factor effects will be documented, their robustness tested, and at last, they will be interpreted. Thus, the analysis, robustness and discussion parts will cover sub-question six through nine. At last, there will be a conclusion based on the findings of this master thesis.

1.4 Delimitations

The problem statement elaborated on in the previous section is broad, and due to that it is necessary to limit the topics of focus. First of all, a vast amount of factors have received academic attention since the early days of factor theory in the 1970's. For reasons of scope, the factors of focus must be defined. Since this thesis focuses both on individual factor effects and models combining several factor effects to best describe the development of different stocks, factors of interest have been chosen by applying two criteria:

1. The individual factor lives up to the MSCI definition of "risk premia factors."[1]
2. The factor is part of a well-recognized equity factor model.

By these two criteria, the following factor models are focused on in the rest of the thesis:

1. The CAPM model of William Sharpe (1964) and John Linter (1965),
2. The three-factor model of Fama and French (1993),
3. The four-factor model of Carhart (1997), and
4. The five-factor model of Fama and French (2015)

As stated, the individual parts of each of these factor models are also an area of focus. These individual factors are the market factor, size, value, momentum, profitability, and investment. The selection of factors is admittedly a bit arbitrary, but it needs to be done in some way.

[1] Factors "which have earned a persistent significant premium over long periods and reflect exposure to sources of systematic risk" (Bender , Briand , Melas, & Subramanian, 2013, s. 7)

Second, the thesis mainly focuses on listed Chinese A-shares. That means that Chinese B shares and non-listed A shares, as well as H shares and mainland overseas listed companies are not extensively covered. However, the overall structure of the Chinese equity market and all of these share classes are described in section 2 on China and the Chinese equity market, where the reasoning for focusing solely on A-shares also is elaborated on. Third, the short history of the Chinese stock market limits the possible period of investigation. Effects in the U.S. stock market have been investigated all the way back to the 1920's, but that is not a possibility in China since the Chinese stock market first was opened in the early 1990's. Several changes and features of the market will also impact the choice of the period of investigation, something that will be further elaborated on in section 5 on data.

Fourth, much of the existing Chinese factor literature uses Chinese databases, as well as MSCI data. The authors of this thesis do not have access to these databases, and the analysis is solely performed with *Datastream* data.

Lastly, market frictions such as transaction costs and shorting barriers are not extensively covered and accommodated for in the analysis. Even though interesting and highly relevant in the Chinese market, it is not a focus of this thesis due to reasons of scope and the exploratory nature of this thesis. Nevertheless, it will be mentioned in different sections as something to bear in mind when relevant. That concludes the introduction part of the thesis, and in the next part the foundation for the analysis will begin with a description of the Chinese equity market.

Section 2 – China and the Chinese equity market

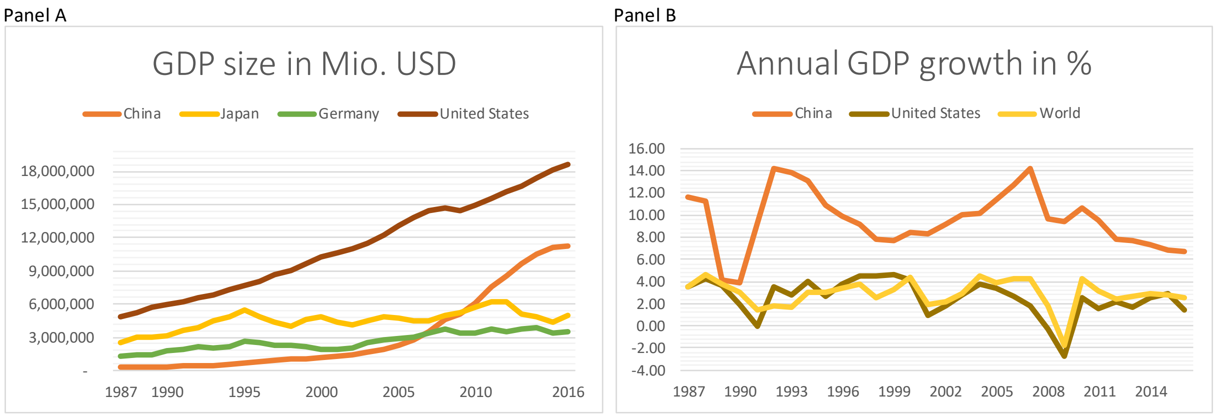
The purpose of this section is to describe the characteristics of the Chinese equity market, as well as the implications the market structure has for the further analysis. This thesis does not concern itself with topics such as the general macroeconomic development of China, but it is nevertheless outlined to a small extent to back up some of the reasons listed for the choice of topic in the introduction and as part of a general introduction. Due to this, the first part of this section gives a general introduction to China and the overall Chinese stock market, before the second part describes characteristics of the market of interest, the A-share market, and what implications its structure has for the further analysis.

2.1 Developments in the Chinese equity market

China has experienced rapid growth during the last three decades. Figure 1 panel A below display the GDP development in China, Japan, Germany, and the U.S. from 1986 to 2016. The underlying numbers from the World Bank can be found in the Appendix. The development over the past 30 years clearly shows the rise of China as an economic superpower. China has gone from being a small fraction of the other three countries in 1987, 9.2 times smaller than Japan, 4.7 times smaller than Germany and 17.8 times smaller than the U.S., to being 2.3 times larger than Japan, 3.2 times larger than Germany, and only 1.7 times smaller than the U.S.

The graph and these numbers illustrate an exceptional growth story, which is further illustrated by the annual GDP growth over the same period displayed in Figure 1 panel B. Also here the underlying numbers from the World Bank can be found in the Appendix. Panel B shows that Chinese GDP has experienced high levels of growth for the entire period, with the growth ranging between 6.7% and 14.2% except for a short drop to approximately 4% in 1989 and 1990. In comparison, the world and U.S. growth are significantly lower for almost the entire period, ranging between -2.8% and 4.7%. The only exception is the years 1989 and 1990. Accordingly, Chinese GDP has grown from 1.6% to 14.8% of world GDP and 7.6% to 49.8% of East Asia & Pacific GDP over the period.

Figure 1 - GDP development overview[2]



Over the same period as the Chinese GDP has experienced extreme growth, the stock market has been evolving rapidly. The market has been fragmented during the whole period investigated above and is still fragmented today. At the most overall level, one can draw the line between stocks listed in mainland China and stocks listed abroad. Listing abroad has been very common for Chinese firms, with the closest link to the Hong Kong Stock Exchange where many firms have been listed as dual listings. That is, they have been listed both on a mainland stock exchange and in Hong Kong. A considerable amount of Chinese firms are also listed in other countries, for example the U.S., Singapore, and

[2] Graphs based on World Bank Data

Germany. As an illustration of the extent of foreign listings, the three giant firms Baidu, Alibaba and Tencent are all listed in the U.S.

One reason to the extent of foreign Chinese listings is claimed to be the strict listing requirements in mainland China, for example the demand for positive earnings in the three years before listing. Allen et al. (2017) find that the post-IPO drop in return on assets is far larger in China than in other countries and that some firms exhaust their resources to meet listing requirements, thus hurting their future growth. Therefore, they conclude that Chinese regulators hurt the stock market by maintaining these regulations and that the firms contributing the most to the country's growth either list abroad or do not list at all. Another reason for a large number of foreign listings could be the complexity of the mainland Chinese share structure. This will be examined in further detail in the next section covering the mainland Chinese stock market.

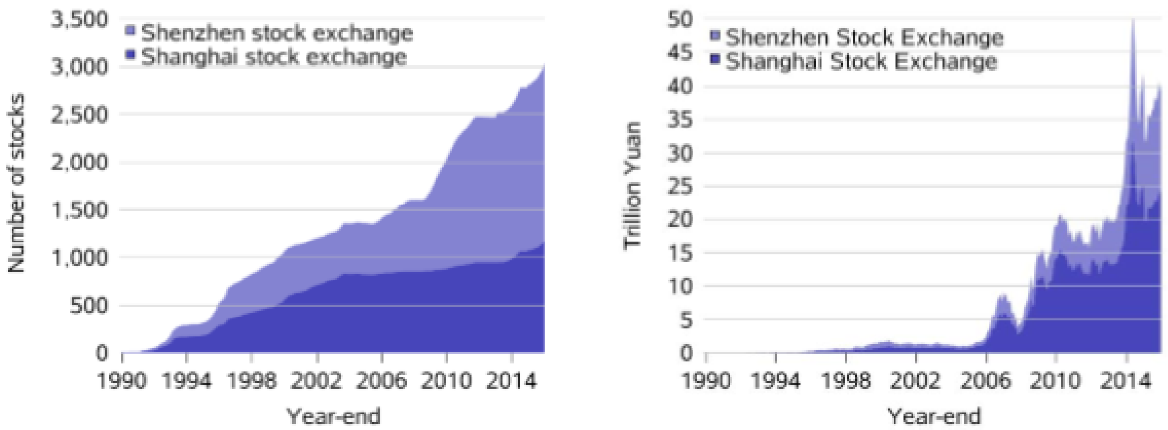
The mainland stock market is a fairly new concept, and as stated above it has been evolving rapidly since its establishment 30 years ago. The Chinese state council announced "Regulations on Deepening Enterprise Reform and Enhancing the Vitality of Enterprises" in December 1986, and after this firms started to issue stocks. As issuing and trading stocks became more and more common, the need for secondary trading was met by the foundation of mainland China's two main stock exchanges: Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE). Both the exchanges were founded in 1990. Hence they have a short history compared to many other stock exchanges. Furthermore, the two exchanges are under the supervision of the China Securities Regulatory Commission (CSRC), which is the main regulator in the Chinese security market. A number of regional security exchanges also operates in the Chinese mainland market, but they have played a minor role both in terms of trading volume and market capitalization[3]. Since its establishment, the Chinese stock market has experienced rapid growth and is today the second largest market in the world by market capitalization[4].

Figure 2 below shows the development in number of firms and floating market capitalization for the two exchanges. The darker blue part shows Shanghai and the lighter blue shows Shenzhen. The Shenzhen part only consists of the part above the Shanghai part, that is, the Chinese stock market consists of around 3 000 listed companies with an overall market cap of around 40 trillion CHY. One can observe from the two graphs that there is a large difference between the two exchanges in number of listed companies, with 1 859 in Shenzhen and only 1 175 in Shanghai by the end of 2016. However, Shanghai has a larger market capitalization, and the differences in number of firms and market cap can be explained by Shenzhen's larger focus on small and medium-sized enterprises, in particular with the

[3] Chinese Capital Market: An Empirical Overview, 2018
[4] Chinese Capital Market: An Empirical Overview, 2018

establishments of the small and medium enterprises board in 2004 and the growth enterprise market in 2009. Hence the firms on the Shanghai Stock Exchange are typically bigger, while Shenzhen Stock Exchange consists of more and smaller firms.

Figure 2 - Size of the Stock Market (1990-2016)[5]



(a) Number of Listed Firms (Monthly) (b) Floating Market Capitalization (Monthly)

Another level of the share structure in China is the different share classes represented by respectively A-shares and B-shares. A-shares are listed in CNY[6], whereas B shares are listed in USD on Shanghai Stock Exchange and Hong Kong dollars on Shenzhen Stock Exchange. The difference between the two share classes comes from the early days of the Chinese stock market when domestic investors only could invest in A-shares, and foreign investors only could invest in B-shares. However, the B-share market was opened to domestic investors in 2001, and the A-share market was at least to some extent opened to foreign investors in 2002. Since then the importance of B-shares has diminished[7], and most literature on the mainland market is only investigating A-shares[8]. Today only 100 listed firms have B-shares, and A-shares account for approximately 96% of the total trading volume[9]. A third share class is H-shares. This share class is used for mainland companies that are listed both on a mainland stock exchange and the Hong Kong Stock Exchange. The shares these companies have listed on the Hong Kong Stock Exchange are called H-shares.

The next level of the Chinese stock market structure to note is the distinction between floating and non-floating A-shares. Floating shares are listed on a stock exchange and are available for the general

---

[5] Reprinted from "Chinese Capital Market: An Empirical Overview, 2018," by Hu, G. X., Pan, J., & Wang, J., 2018, NBER WORKING PAPER SERIES, NBER Working Paper No. 24346, page 6. Copyright 2018 by Hu, G. X., Pan, J., & Wang, J.

[6] The Chinese currency is called Renminbi (RMB), and is denominated in Yuan (CNY).

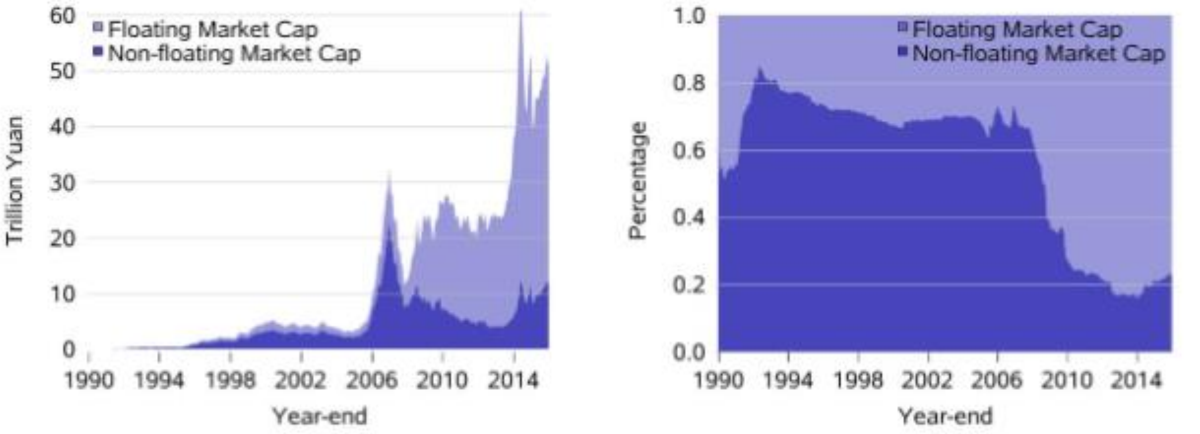
[7] https://www.ft.com/content/254b3b6e-5a2a-11e2-a02e-00144feab49a

[8] See for example Wang and Xu (2004), Xu and Zhang (2014), and Hu et al. (2018).

[9] Chinese Capital Market: An Empirical Overview, 2018

public, whereas non-floating shares are neither listed nor traded on an exchange. The latter share class is only traded in-over-the counter markets. This system is called the “split-share structure,” and the relationship between the two share classes is referred to as the floating ratio. The structure stems from the early days of the Chinese stock market in the 1980’s when the Chinese government introduced and actively promoted non-state ownership in state-owned firms. Gradually, more and more non-floating shares have been floated, and thereby they have become available to the general public. Figure 3 below shows the development in floating ratio from 1990 to 2016, and it shows that floating shares have overtaken non-floating shares as the dominating ownership class in the mainland equity market.

Figure 3 - Developments in the floating ratio[10]



(a) Market Capitalization (Monthly) (b) Share in Market Capitalization (Monthly)

Up until the launch of the “Stock Connect Program,” floating A-shares could only be traded by domestic individual or institutional investors, financial intermediaries and financial service providers, and qualified foreign institutional investors (QFII). The access for foreign investors was limited due to the need for a license and a designated quota to invest. The “Stock Connect” program is in that respect a new channel that eases restrictions and opens up the Chinese market to a larger extent. The pilot program started as a collaboration between Shanghai Stock Exchange and Hong Kong Stock Exchange in November 2014 and was also extended to Shenzhen Stock Exchange in late 2016. Shortly, “Stock Connect” makes it possible for international investors to invest in selected stocks in the China A-share market in a less restrictive manner than earlier through the Hong Kong Stock Exchange, and for mainland Chinese investors to invest in selected listings on the Hong Kong Stock Exchange.

[10] Reprinted from “Chinese Capital Market: An Empirical Overview, 2018,” by Hu, G. X., Pan, J., & Wang, J., 2018, NBER WORKING PAPER SERIES, NBER Working Paper No. 24346, page 11. Copyright 2018 by Hu, G. X., Pan, J., & Wang, J.

Many regard the "Stock Connect" program as a further step in the on-going process of integrating the Chinese market into global markets[11]. It was also one of the main reasons listed when MSCI published that China A-shares would be included in their Emerging Markets Index[12]. Specifically, the program makes Chinese shares more accessible by allowing investors to invest in the A-share market without an individually assigned quota or license, and the overall market access quota is also significantly raised. "Stock Connect" has therefore been perceived by global markets as a strong signal of the Chinese willingness to open up more.

As stated in section 1.4, this thesis mainly investigates the A-share market. A large part of the motivation for the thesis is the isolated nature of the Chinese market, and due to this H-shares and other foreign listings are overlooked in the further analysis. In addition, B-shares are not an area of focus because of the minor importance they have in the overall market and the focus of existing literature on A-shares. At last, non-traded A-shares are not an area of focus due to the lack of public information on these shares. This part has outlined the overall equity structure in China, and the thesis narrows in on China A-shares in the next part as it is the focus of the further analysis.

2.2 China A-share characteristics

MSCI characterizes China as an emerging market[13], and stock return properties align well with earlier observations of emerging stock market returns[14]. Carpenter, Lu and Whitelaw (2018) report high normal and risk-adjusted excess returns, high volatility, and low correlation with three big developed markets (U.S., Europe, and Japan) for mainland China over the period from 1995 to 2016. These properties are attractive for asset managers as they provide diversification benefits. Other characteristics of the Chinese A-share market worth noting is that the market to a large extent is held by Chinese retail investors[15] and that it has a very high turnover ratio[16].

There are furthermore several peculiar regulations in the mainland Chinese market that might have implications for the behavior of Chinese stocks. First of all, a daily price limit has been effective since 26th of December 2006, which implies that any single trading day, the price of a stock can maximum increase or decrease 10% from the closing price of the previous day. It should also be noted that there are several exemptions from this rule[17]. Secondly, short selling is difficult. It has been legal to short sell

---

11 See for example (Ma, 2014)

12 https://www.msci.com/eqb/pressreleases/archive/2017_Market_Classification_Announcement_Press_Release_FINAL.pdf

13 https://www.msci.com/market-classification

14 See for example Harvey (1995) and Fama and French (1998)

15 Eun and Huang (2007) and Cheung et al. (2015) both report that retail investors make up 90% of the investor base.

16 Eun and Huang (2007) report annual turnover ranging from 117% to 1048% during the 14-year period they measure, with an average of 483%.

17 See Chinese Capital Market: An Empirical Overview (2018) for list of exemptions.

since 2006, but it is difficult in practice[18]. Thirdly, what is referred to as the T+1 rule is that stocks bought on day t are settled on day t+1. That also means that stocks bought on day t at earliest can be sold at day t+1. All stocks on Shanghai and Shenzhen Stock Exchange are subject to this rule.

Fourth, there are some taxation rules worth noting in the Chinese market. The first thing to note is that capital gains and dividend income are taxed unequally. Capital gains are not taxed at all, while dividend income is taxed. The dividend tax rate was initially 20% but was reduced to 10% in June 2005. Dividend tax regulations were changed again in 2016, when the taxation rate was lowered to 5% for holding periods longer than one year, remained at 10% for holding periods between one month and one year, and was increased to 20% for holding periods shorter than one month. Those are the rules that currently apply. They were introduced to limit speculation on stocks with high dividend income. A second peculiar tax feature is a transaction tax referred to as the stamp tax. The stamp tax is currently 0.1% on the sell side of a trade, but it has previously been charged on both sell and buy side, and the rate level has been changed several times since the foundation of the mainland Chinese stock exchanges.[19]

Fifth and last, firms with what the Shanghai and Shenzhen Stock Exchange characterize as financial abnormality receives a special treatment status ("ST" status). The prefix "ST" is then added to its stock name, and for ST firms the daily price limit is only half of the normal price limit. Given the difficult listing process, the listing itself is valuable, and firms are rarely delisted. Allen et al. (2017) find that Chinese firms experience similar drops in performance before receiving ST status as U.S. firms do in the five years prior to their delisting. The difference is that the U.S. firms disappear from the market at this point, whereas Chinese listings either recover or are often taken over by a private firm seeking public listing. There are furthermore different kinds of ST status, depending on the severity of the firm's financial problems[20].

This section on China and the Chinese equity market first outlined the overall structure of the equity market and explained the reasons for looking at the specific part of the market this thesis is looking at, namely the tradable A-share market, before characteristics and peculiarities of the A-share market was further investigated. How the characteristics and peculiarities are accommodated for in the analysis will be further discussed in the data and methodology parts later, based on the information given in this part and the literature review in the next part.

[18] Chinese Capital Market: An Empirical Overview (2018)

[19] See Chinese Capital Market: An Empirical Overview (2018) for specific details on the stamp tax rate level in different periods.

[20] Firms with ST status have experienced losses for two consecutive years, firms with ST* status for three consecutive years, firms with SST status for two consecutive years and have not completed the stock split-structure reform and firms with S*ST status for three consecutive years and have not completed the stock split-structure reform.

Section 3 – Literature review

The purpose of this thesis is to investigate factor model explanations of equity returns in the Chinese A-share market, and the literature review is therefore organized in a general part on the framework for equity factor models and a more specific part on this type of research in China. The general part is further divided into a fundamental factor research part to outline the framework factor models operate within, a part on the main factors investigated in this paper, and a part on some of the critique factor models have faced. The literature review is extensive, but the extent of it is necessary to build a solid framework for a topic with a wide reach.

3.1 General introduction

The first part of the literature review introduces the general framework factor models operate within before the specific parts of each factor model are further examined by looking at the single factors they consist of. At last, a section with critique of the factor models introduced is included. As stated in section 1.4, this thesis focuses on the CAPM of Sharpe (1964) and Linter (1965), the three-factor model of Fama and French (1993), the four-factor model of Carhart (1997), and the five-factor model of Fama and French (2015). These models are chosen because they are well recognized in modern equity research. For simplicity, they will be referred to as the CAPM, the three-factor model, the four-factor model, and the five-factor model. Many of the individual factors are common for several of the listed factor models, and the list of factors is therefore not as extensive as it might appear judging by the number of different models.

3.2 Framework for factor models:

CAPM

The framework for factor models comes from early works on The Capital Asset Pricing Model (*CAPM*)[21]. The *CAPM* was developed by John Linter (1965) and William Sharpe (1964) in the 1960´s to estimate expected returns of assets and still appears to be one of the most frequently used statistical models in modern finance. Despite a range of criticism, empirical work and several (successful or unsuccessful) attempts to reject the *CAPM*; it has been the main starting point for later asset pricing literature[22]. The

[21] The classical *CAPM* assumption: $E(\tilde{R}_j) = \gamma_1 \beta_j$ where $\gamma_1 = E(\tilde{R}_M)$ is the expected excess return of the market

[22] The *CAPM* formula with intersect: $R_i - R_f = \alpha + \beta(R_M - R_f)$. In a beta – expected return universe a straight line between the rf rate, located on the y-axis (expected return) and the market portfolio (with a market return and a beta equal to 1) can be drawn. This line is called the security market line (*SML*). If the *CAPM* holds all the assets will be located on the SML and alpha will be equal to zero.

model itself is based on Markowitz mean-variance efficient theory (1952), where investors choose their portfolios along the efficiency frontier to achieve an optimal allocation of return and risk. It assumes that investors can invest in a risk-free rate, all investors have mean-variance preferences, and all investors have homogenous awareness of risk-free rate and mean-variance preferences. If these three conditions hold and the market is in equilibrium, the tangency portfolio will be equal to the market portfolio of all risky assets. In addition, each investor holds a combination of the market portfolio and the risk-free rate, and due to that, the linear *CAPM* relation between expected returns and market risk holds[23].

The *CAPM* is important for further academic work because it links expected returns from holding a specific asset to overall market returns. The beta (the exposure to the market factor or later called the factor loading of the market factor) measures how much the single asset covaries with the market and is due to that a proxy for the risk of an asset. Following Sharpe-Linter's work the *CAPM* is the market exposure (β_M) that helps to explain the future return of an asset. The *CAPM* itself is set-up as a risk-based model where higher risk is related to higher future return. Effectively, this means that investors are compensated with higher returns for taking on higher risk. The *CAPM* is a framework that focuses on the systematic risk[24] of stocks rather than on the idiosyncratic risk[25]. Later factor models are largely based on the risk assumption and the associated covariance of an underlying asset with a factor.

Performing an empirical study on the *CAPM*, Black, Jensen and Scholes (1972) are able to show that the classical *CPAM* assumption does not hold due to various reasons. They provide evidence for an existing persistent deviation from the *SMB*. This deviation component is expressed as alpha[26]. They define the classical *CAPM* to hold under the sub-condition that all alphas (alpha_j) are equal to zero. However, the empirical *CAPM* has a flatter relationship between returns and risk than expected by the conceptual *CAPM*[27] (Black, Jensen, & Scholes, 1972).

The *CAPM* is mostly tested in an intertemporal framework, where assets are assumed to have constant expected returns and market exposure (*ß*), and the average market risk premium stays constant over time. Investors are required to choose the optimal mean-variance portfolio at the start of each period. For investors with multi-period investment horizons, the previous assumption of constant expected

23 The conceptual *CAPM* equation $\mathrm{E}[r_i] - r_f = \beta_i(\mathrm{E}[\mathrm{r}_m] - r_f)$
24 Systematic risk arises from underlying behaviour of the market and affects all equities. It is typically non-diversifiable and therefore omnipresent.
25 Idiosyncratic risk is the stock specific risk that might only apply to the individual stock due to certain underlying characteristics of a company. This company specific risk will disappear as soon as well diversified portfolios are formed.
26 Where $\alpha_j = E(\tilde{R}_j) - E(\tilde{R}_M)\beta_j$
27 Further evidence on this issue is provided by Fama and MacBeth (1973)

returns will not be true (Merton, 1973). As investors dislike future uncertainty, they invest in long-term low-risk assets, which will drive down future returns of such assets. Reversely the price of more risky assets will be driven down, and their returns will rise. Merton (1973) derives this in his intertemporal *CAPM* framework and concludes that expected returns on risky assets will differ from risk-free assets even if these assets have a market beta of zero (no market exposure). Thus, this leads to deviations from the conceptual one-period *CAPM*.

Another issue for the *CAPM* to hold was formalized by Roll (1977). He argues that the results of the *CAPM* and other asset pricing models might never be correct as these models are based on the market of all risky assets, which is never entirely observable[28].

APT

In his influential paper on the Arbitrage Pricing Theory (*APT*) Ross (1976) proposes an alternative to the mean-variance based *CAPM* for asset price estimation. The *APT* does not require the restrictive "homogeneity of anticipation of the mean-variance theory" (Ross, 1976, *p.355*). As mentioned before, this has been a major challenge for the *CAPM,* especially when using the *CAPM* in an intertemporal framework. Different risk aversion levels and differently endowed individuals will always have different preferences towards their asset allocation choices. Ross (1976) assumes a world where prices are in equilibrium. As soon as there is an arbitrage opportunity arising, it will be exploited by some investors. The higher demand for these arbitrage investments drives up their price until prices are in equilibrium again.

One of the major differences between the *CAPM* and the *APT* is the presence of multiple price-influencing factors. These factors are called state variables and link overall economic developments with asset prices. Ross (1976) assumes a K-factor model, whereas the *CAPM* assumes a one-factor model (only the market factor matters in the *CAPM*). For the *APT* to hold, returns are described by a K-factor model, arbitrage opportunities disappear through efficient pricing, and investors can diversify the idiosyncratic risk of single stocks away through portfolio formation (Ross, 1976). If these assumptions hold, expected returns are approximately described by a combination of factors. Alpha, the outperformance of the expected returns over the factor described returns should be (close) to zero. To quantify the effect of changes in underlying state variables/factor behavior and how they interact with asset prices, quantifiable, investable portfolios are built. These portfolios mimic the

[28] In general, especially non-listed trades like OTC trades of securities, real estate and other assets are unobserved and can hardly be included into a market estimation

behavior of its underlying factors (Ross, 1976). In modern asset pricing theory factor mimicking portfolios are a frequently used tool when aiming to explain factor returns.

In their paper, Chen, Roll and Ross (1986) define the *APT* model as a five-factor model of underlying macroeconomic factors. They show that the exposure of an asset to state variables/factors describe asset prices. The factors influence either the discount rate of future dividends or future dividends/ earnings itself or both together[29]. As statistical significant influential state variables, Chen et al. (1986) find five factors describing asset prices: long and short run government bond spread, changes in expected inflation, changes in unexpected inflation, the growth of industrial production, and corporate bond yield spread.

To sum up, the *CAPM* is essentially a one-factor model that seeks to describe stock returns by using the market risk and a stocks exposure to the market risk. Its assumptions are precisely defined and rather rigid. The *APT* model builds on similar assumptions appending the risk factor idea with several additional risk factors. Its assumptions are less rigid, which leads to an approximate validity of the equation. Later literature in the 80´s and 90´s heavily bases on these two fundamental models. For example, Fama and French (1992, 1993) develop their models mainly in-line with the *APT* model. Most of these models assume a positive causal relationship between risk and return. However, it is necessary to point out that the mechanisms of the *APT* also work in a more fundamental, non-causal relationship between risk and return.

Expanding the CAPM

Not only the theoretical models covered in the previous section challenged the CAPM, but the simple and linear relationship of covariance with the market as the sole measure of risk also faced several empirical challenges. The relation was challenged by the size effect (*Size*)[30] of Banz (1981), the leverage effect of Bhandari (1988), the earnings-to-price ratio effect of Ball (1978) and Basu (1983), as well as the book to market value of equity (*B/M*) ratio described in papers by Stattman (1980), Rosenberg, Reid and Lanstein (1985) and Chan, Hamao, and Lakonishok (1991). All of these works document high returns that cannot be explained by the *CAPM*. It was with this background Fama and French wrote their famous 1992/1993 paper on factors that affect the cross-section of expected stock returns.

[29] Chen, Ross and Roll (1986) assume that the price of an asset is defined as $p = \frac{E(c)}{k}$, effectively expected dividend cash-flow divided by the discount factor k

[30] The variable Size that is further referred to is based on the market value of equity of a company

Three-Factor Model

Fama and French (1992) examine the joint role of the *size* and *B/M* variables investigated in the papers mentioned above, in addition to market beta (*beta*). As earlier mentioned, they base their model on the *APT* framework of Ross (1976), and among other things, they find that *beta* does not help to explain cross-sectional variation in stock returns in their sample. The relation between *beta* and average return that is unrelated to *Size* is found to be flat. They also find that the *Size* and *B/M* variables to a large extent explain the variations in cross-sectional returns and that the combination of *B/M* and *Size* absorbs the leverage and earnings-to-price effect. They further state that their tests and results impose a rational asset pricing framework with the *B/M* and the *Size* factors as proxies for underlying risk factors that affect stock returns, as long as assets are rationally priced. That is, they draw the line to and build on the models of Merton (1973) and Ross (1976).

From the return patterns, they observe they try to work backward to the underlying economic risk factors. Fama and French (1992) suggest future research to focus on the relation between their two factors, *B/M* and *Size*, and the underlying economic risk factors they are supposed to proxy for. Exactly that is what they do in the following years. Building on the *APT* framework, their overall picture is that if average return relations are due to rational pricing, two conditions must be fulfilled:

1. There have to be common risk factors in returns related to *Size* and *B/M*
2. *Size* and *B/M* effects in returns should be explained by earnings

Fama and French (1993) investigate the first condition, and they consider their findings as evidence of the existence of common risk factors in returns associated with *Size* and *B/M* ratio. They investigate the time variations in stock (and bond) returns by forming portfolios constructed to mimic risk factors. Again they find that *Size* and *B/M* explain cross-sectional differences in stock returns. However, the large difference between the risk-free rate and average stock returns cannot be explained solely by these two factors. That makes them add the market factor to the other two factors in their regression, thus completing the famous three-factor model that they show to explain much of the variation observed in stock returns.

The second condition from above is investigated in Fama and French (1995). The links between their risk factors and economic fundamentals are explored by asking if the behavior of stock prices associated with *Size* and *B/M* reflect the behavior of earnings. They document that common factors in returns mirror common factors in earnings, and therefore suggest that the market-, *Size* - and *B/M* effects in earnings are the source of the same effects in returns. Thus these factors proxy for underlying risk factors since they also are present in earnings. Furthermore, they find that market and *Size* effects in earnings also explain the same effects in returns, while they fail to find the same for the *B/M* effect.

Four-Factor Model

Some years after the influential initial Fama and French papers, Carhart (1997) introduces a four-factor model by appending a momentum factor to the three-factor model of Fama and French (1993). He constructs the whole model in the spirit of Fama and French, in the sense that the four risk mimicking portfolios are supposed to proxy for risk factors. However, Carhart does not specify the risks his factors proxy for. The model has nevertheless been influential since its introduction, and for example Fama and French (2012) investigate the momentum factor together with their three original risk factors. They find strong evidence of momentum patterns in returns in North America, Europe, and the Asia Pacific. The only area they fail to detect such patterns is Japan.

Five-Factor Model

In recent years, Fama and French (2015) have extended their three-factor model with a profitability factor and an investment factor, thus formalizing what often is referred to as the Fama and French five-factor model. In this paper, they extend the methodology developed in Fama and French (1993) to investigate profitability and investment anomalies that have been shown to cause problems for the three-factor model by among others Novy-Marx (2013) and Titman, Wei and Xie (2004). They are able to show that the five-factor model better captures the variability of stock returns in the U.S. market than their earlier three-factor model, a result that naturally has received broad attention. Furthermore, Fama and French (2017) test the same five-factor model in global markets. They find the five-factor model to be superior in Europe, North America and the Asia Pacific, while they only observe a clear *B/M* pattern in Japan. However, they also find the investment factor to be redundant in Europe.

That concludes the general part of the literature review. This section has introduced the general framework equity factor models operate within and the four models this thesis mainly focuses on: the CAPM, the three-factor model, the four-factor model, and the five-factor model. The three latter build on the general APT framework introduced by Ross (1976). The influential methodology applied in Fama and French (1993) and further developed in papers like Fama and French (2015) is the main reference point in this thesis. The next section proceeds with the introduction to the individual factors that make up the different factor models of interest.

3.3 Individual factors

The factors that are introduced in this section are all a part of frequently used factor models. As previously stated, the focus of this thesis is the CAPM, the three-factor model, the four-factor model, and the five-factor model. Factors that will be further elaborated on in this section are therefore the size and the value factors (three-factor model), the momentum factor (four-factor model), and the profitability and investment factors (five-factor model). The equity premium is considered adequately elaborated on in the general framework for factor models and is therefore not elaborated on in the specific factor part. It needs to be pointed out that several explanations of factor returns exist, typically both risk-based and behavioral explanations. This thesis reviews papers from the two competing views to some extent. The first factor to be introduced is the value factor.

Value

Academic research on the value premium has a long history, reaching back all the way to *Security Analysis* by Graham and Dodd (1934). It exists a wide range of literature on this premium, and it is not by any means fully reviewed here. That is beyond the scope of this thesis. Nevertheless, the section reviews some of the most recognized papers to highlight the most important features of the value premium.

Value investors compare a measure of a stock's fundamental value to the stock's prevailing market value, and systematically invest in stocks with a high fundamental value relative to market value. The idea is that the price of a stock equals expected future cash flows divided by the expected return. This is illustrated in the equation below, where P_0 is the current market value, $E(CF_t)$ is expected future cash flows at different times t, and r is the expected stock return.

$$P_0 = \sum_{t=1}^{\infty} \frac{E(CF_t)}{(1+r)^t}$$

By isolating the expected return on the left-hand side of the equation, expected return equals expected future cash flows divided by the current price of the stock. The perfect measure of fundamental value would be the actual future cash flows, but this is not achievable at a previous point in time. In this framework, several other scaling variables of price have been shown to be useful for predicting future stock returns, such as the book value of equity, earnings, dividends, and cash flows. Fama and French (1992), inspired by earlier observations of the *B/M* anomaly by Stattman (1980), Rosenberg, Reid and

Lanstein (1985) and Chan, Hamao, and Lakonishok (1991), observe a strong positive relation between the *B/M* ratio and average excess returns.

Fama and French (1992) propose two interpretations of the *B/M* variable. The first and most intuitive interpretation they offer is that the market judges firms with a high *B/M* ratio, i.e., a low price relative to its book value to have poor earnings prospects compared to firms with a low *B/M* ratio. Therefore, they suggest that the *B/M* ratio captures a relative distress risk previously investigated for the size anomaly by Chan and Chen (1991). Due to this distress risk firms with a high *B/M* ratio earn a premium.

Secondly, they run regressions of excess returns on two leverage ratio variables, market leverage, and book leverage, and find that the slopes on the two leverage variables have opposite signs but are close in absolute value. They observe that these opposite roles to a large extent are captured by the *B/M* variable. Hence, they conclude that the *B/M* ratio can be interpreted as a market imposed leverage effect, which is captured by the difference between market leverage and book leverage[31]. A high *B/M* ratio says that a firm's market leverage is high relative to its book leverage due to the market's assessment of the firm's earnings prospects as poor. In addition to the leverage effects *B/M* is found to capture, Fama and French (1992) also find that most of the previously documented positive relation between the earnings-to-price ratio (*E/P*) and average excess returns[32] is due to the positive correlation between *E/P* and *B/M*.

Fama and French (1995) find that *B/M* and size effects in returns mirror the same effects in earnings, which suggest that the effects are caused by fundamental risk factors. This suggestion is further supported by the later findings in Fama and French (1998), where they show that value stocks have higher returns than growth stocks in 12 out of 13 markets investigated. That is, the value premium is present across markets, which suggests a global risk factor. However, Fama and French (1995) fail to find causality in the relation between the *B/M* effect in earnings and the same effect in returns.

Zhang (2005) points out that there is an inconsistency between the risk-based explanation of the value premium and the common perception of growth firms as riskier than value firms. Therefore, he tries to shed light on this apparent puzzle. This is done by arguing that value firms, in fact, are riskier than growth firms due to costly reversibility and countercyclical price of risk. That is, it is harder to reduce current assets in bad times than it is to expand in good times. Firms with a high amount of current assets (high *B/M*) will be harder hit because it is costly to reduce assets and they will struggle with

[31] Market leverage (ln(A/ME)) has a positive sign and book leverage (ln(A/BE)) has a negative sign, and book to market equity can therefore be described by the relation ln(BE/ME) = ln(A/ME) – ln(A/BE)

[32] See for example Ball (1978) and Basu (1983)

more unproductive capital, while growth firms do not have this issue. At the same time, growth firms will invest more in good times to take advantage of good economic conditions, and due to that be less profitable in good times than value firms. These two effects sum up to make value firms more exposed to both good and bad economic conditions and thus carry a significant risk premium.

Asness, Moskowitz and Pedersen (2013) find value strategies to have worked across global markets and asset classes. They also find value strategies across markets and asset classes to be positively correlated, which they suggest indicates evidence of a global systematic risk factor. This is consistent with the risk-based framework Fama and French argue for. However, they point out that the value premia across asset classes are inconsistent with rational asset pricing theories explaining the value premium with investment risk and growth options. That makes it a problem for the prevailing rational risk-premium explanations.

Lakonishok, Shleifer and Vishny (*LVS*) (1994) present an alternative to the underlying risk explanation of the value premium. They take a behavioral view focusing on the underlying reasons to why value strategies outperform growth strategies. According to their findings value strategies work/pay a premium because they extrapolate mispricing in the market. This mispricing stems from the larger attractiveness of growth shares (shares issued by fast-growing companies) compared to value shares (shares issued by companies with slower growth). Growth stocks appear more "glamorous" especially to naïve investors, and therefore they face higher demand. *LVS* (1994) state that this happens although naïve investors are unaware of/not interested in the underlying reasons for a company's performance. This increased demand drives up the price of growth stocks, thereby lowering their expected future returns. Moreover, they conclude that value strategies do not deliver premia due to higher inherent risk compared to growth strategies. They consider both strategies equally risky. Apart from the possible existence of priced factors, the estimated magnitude of return premia on factors might be excessively high and the correlation with underlying macroeconomic factors are too low to represent systematic risk compensation.

To sum up, solid value premia have been found across markets and asset classes. Fama and French (1993) explain the value premium by the distress risk of Chen and Chan (1991), Zhang (2005) attributes it to costly reversibility and countercyclical price of risk, and a behavioral view presented by LSV suggests that the lower returns of "glamour stocks" are caused by their attractiveness compared to value stocks. Regardless of the reason for the value premium, the academic awareness of it traces back all the way to Security Analysis by Graham and Dodd (1934), and it still appears to be one of the most solid premia today. Next, the thesis moves on to the size premium.

Size

Banz (1981) was one of the first to document the "*size* effect" for NYSE shares and is one of the papers Fama and French base their initial research on. Specifically, Banz finds that smaller firms on average have higher risk-adjusted returns than larger firms. The *size* effect is believed to have disappeared after the discovery of it in the early 1980's since there is no significant size effect in U.S. stock data after the mid-1980's[33]. In international stock returns, for example Fama and French (2012) find no significant size effect either. The literature on reasons for the *size* effect is therefore not as extensive as for the value premium but is rather focused on possible reasons for the disappearance of the *size* effect. Nevertheless, several reasons for the *size* premium have been proposed, and some of these will be covered next. At the end of the section, some explanations for the disappearance of the *size* effect are briefly touched upon.

Banz (1981) does not investigate whether it is the size of a firm that is the underlying risk factor, or if it proxies for other factors through shared correlation. However, he offers two possible explanations for the case that size is the underlying risk. The first is that the source of the premium could be rooted in mergers. That is, larger firms are able to pay a premium for smaller firms since they can discount the same cash flows at a lower discount rate. Secondly, he suggests a model built by Klein and Bawa (1977) as a possible explanation. They argue that insufficient information about a security is a reason not to hold it due to estimation risk. It is likely that there is less available information on smaller firms, and smaller firms therefore provide higher returns due to the additional estimation risk.

Fama and French (1992) find that the positive relation between *E/P* and *B/M* previously described for value does not capture the high excess average return of firms with negative earnings. This relation normally postulates that a high *E/P* is associated with a high *B/M* and future excess return, but this relation is disturbed by the high returns of firms with negative earnings. They find the size factor useful for explaining the high returns of negative earnings firms, as adding the size variable to the three-factor model removes all the explanatory power of their negative earnings dummy.

Fama and French (1993) document a negative relationship between size and average returns. They also observe that small firms do not participate in the stock market boom of the late 1980s, and therefore suggest that the premium small firms seem to be paying is due to a common risk factor for small firms that leads to prolonged small firm earnings depressions that largely passes by bigger firms.

[33] Asset Management: A Systematic Approach to Factor Investing, Ang (2014)

In their 1995 paper, Fama and French document that the size effect in earnings also explains the same effect in returns, and therefore conclude that it is a proxy for underlying risk factors. However, they do not specify the underlying factors size is supposed to proxy for.

Chan and Chen (1991) argue that the reason small firms earn a premium compared to big firms is not their size in isolation, but rather that the portfolio of small firms are more populated with what they call marginal firms. They call these firms marginal firms because they have low production efficiency and high financial leverage, and are therefore more sensitive to changes in the economy. That is, their systematic risk is large, and they pay a premium relative to firms carrying less systematic risk.

As mentioned initially, the disappearance of the *size* effect has also received some academic attention. One explanation for the disappearance of the size effect proposed by among others Fischer Black (1993), is that there never was any *size* effect. This effect just appeared to be there by chance in the sample periods used in initial literature. The investigation of the *size* effect is therefore according to this view a result of a type 1 error. That is, a statistical relationship that is only statistically significant by chance is accepted. Another explanation is that the abnormally high returns of small stocks were overexploited after the initial discovery. That is, small stocks were subject to larger buying pressure, thus pushing their current prices up and future returns down. China represents a particularly interesting case in regards to these two opposing views, as it has been a largely isolated market with little influence from the outside.

To sum up, Banz (1981) suggests that the observed *size* premium can be caused by mergers and that big firms are able to discount cash flows at a lower rate than small firms, or because of estimation risk due to insufficient information on small firms. Fama and French (1993) suggest that it can be caused by prolonged earnings depressions that pass by other types of firms, while Chan and Chen (1991) proposed that the reason to the *size* premium is the high proportion of marginal firms among small firms. At last, the believed disappearance of the *size* factor is especially interesting to investigate in the isolated Chinese market. Next section covers the momentum factor.

Momentum

Jegadeesh and Titman (1993) were one of the first successfully to investigate momentum strategies in newer literature. At the time they wrote their initial paper on momentum, they were motivated by the lack of academic work on the topic, at the same time as practitioners seemingly were successfully pursuing momentum strategies. The academic work on the topic only managed to document successful reversal strategies, but Jegadeesh and Titman observed that there was a discrepancy between the time horizon academic work and practitioners had. Whereas academics used either very short horizons (one week or one month) or very long horizons (three to five years), anecdotal evidence suggested that practitioners followed momentum strategies based on price movements over the last three to twelve months. Their suspicions turned out to be spot on, as they document significant profits for each of the momentum strategies they investigate. These strategies make the investment decision based on price movements over the last 3- to 12-month period and have holding periods ranging from 3 to 12 months. Furthermore, Jegadeesh and Titman suggest that the profits of the momentum strategies they investigate are not due to systematic risk. Their evidence is consistent with a delayed price reaction to firm-specific information explanation.

Another influential paper on the momentum factor is presented by Carhart (1997), as mentioned in the general part. His main result is that almost all mutual fund returns are attributable to common investment strategies and transaction costs. While this perhaps not is of great relevance to this paper, several other aspects of his work are. First, he concludes that transaction costs consume the profits of following a momentum strategy. This is important to note for the practical use of the strategy. Second, he constructs the momentum factor based on price movements over the last eleven months lagged one month. That is, the last 12 months skipping the last month. This way of constructing the medium-term momentum factor has gained broad acceptance. Third, as mentioned in the general section, Carhart introduces a four-factor model by adding the momentum factor to the three-factor model of Fama and French (1993).

Jegadeesh and Titman (2001) later update their results from 1993 on the momentum strategy. They do this partly to address criticism concerning that their previous results are caused by data mining, and partly to review and assess the different academic explanations for the profitability of the strategy. First, they perform out-of-sample tests to address the issue of data mining. The tests on data for the eight years after their initial paper show "remarkably similar" results compared to the results obtained from the initial sample, which mitigates the concerns of data mining. In fact, the similar results also provide evidence of persistence in the detected return patterns. This is noteworthy since other effects, such as the *size* effect of Banz (1981) previously covered has weakened after the academic detection and initial publications on the topic.

Second, Jegadeesh and Titman test the validity of two opposing explanations for the profitability of the momentum strategy. As usual, the opposing views are behavioral and risk-based. Numerous papers have been written on the former explanation[34], whereas Jegadeesh and Titman reference Conrad and Kaul (1998) as an example of the latter. The behavioral view roots in the explanation of the contrarian strategy of De Bondt and Thaler (1985). They show that stocks that have performed poorly over the previous three to five years perform better than stocks that performed well over the same period, and explain this with the overreaction of stock prices to information. Since the stock price has overreacted to the information, a reversal movement to the fundamental value will occur as the information is correctly interpreted. According to the behavioral explanation of momentum, the strategy is successful over a medium horizon due to this delayed overreaction that the contrarian strategy later trades on, as well as an initial underreaction. Conrad and Kaul (1998) argue that cross-sectional variations in expected stock returns rather than predictable time-series variations cause momentum profits. That is, risk characteristics of firms are what determine momentum returns.

The implications of these two views oppose each other. According to the behavioral view, the stock price should correct down after the period of overreaction, while the stock should continue to deliver higher returns according to the risk-based view. Jegadeesh and Titman (2001) therefore assess these views by examining stock returns over a 60-month period. They find significant positive returns during the first 12 months after the portfolio formation date, followed by negative cumulative returns from month 13 to month 60. That is, the stock price corrects down after the momentum period. This is in line with the behavioral view, while it conflicts with the risk-based view. However, they do insist on a cautious interpretation of their results. Their momentum returns are robust, but the contrarian returns seem to be dependent on the sample and period of investigation.

There have been later attempts to document risk as the explanation to momentum returns. Griffin, Ji and Martin (2003) investigate whether macroeconomic risk can explain the momentum profits in 40 different countries, but fail to document any such explanation. Contrary to what they set out to do, they document large and statistically significant momentum returns around the globe in both good and bad economic environments and observe reversal patterns of these returns over the next one to five years. Their attempt to document underlying risk factors for the momentum strategy thus fails, as the reversal pattern rather supports the behavioral explanation. Other attempts to document risk as the driver of momentum profits have been more successful, with for example Johnson (2002)

[34] For example Barberis, Schleifer and Vishny (1998), Daniel, Hirshleifer and Subrahmanyam (1998) and Hong and Stein (1999)

connecting it to firm growth rate risk, Sagi and Seasholes (2007) connecting it to revenues costs and growth options, and Liu and Zhang (2008) connecting it to the growth rate of industrial production.

As described for value, the findings of profitable and correlated momentum strategies across global markets and asset classes by Asness, Moskowitz and Pedersen (2013) support the explanation of underlying risk as the source of momentum profits. At the same time, it challenges the previously mentioned growth based rational explanations.

Independent of the reason for it, momentum patterns in returns have been detected in large parts of the world. Jegadeesh and Titman (1993) first documented momentum patterns in returns, and it has become a highly recognized factor premium after this due to significant findings in different markets. Carhart (1997) created the four-factor model by appending the momentum factor to the three-factor model of Fama and French (2015), and this paper will investigate momentum in the Chinese marked both as a single factor and as a part of Carhart's four-factor model.

Investment and profitability

Profitability and investment are the two factors Fama and French (2015) add to their initial three-factor model from Fama and French (1993). The two factors are added partly due to the economic reasoning backing up these two factors, and partly due to the challenges these two variables cause the three-factor model. Fama and French (2015) state that they are aware of the fact that many other variables also have been shown to cause problems for the three-factor model, but these two variables are natural choices for them because of the economic interpretation they have. Specifically, they start with the Miller and Modigliani (*MM*) (1961) equation,

$$M_t = \sum_{\tau=1}^{\infty} \frac{E(Y_{t+\tau} - dB_{t+\tau})}{(1+r)^{\tau}}$$

Where $Y_{t+\tau}$ is total earnings for the period from t to τ, $dB_{t+\tau}$ = $B_{t+\tau}$ - $B_{t+\tau-1}$ is change in book value of equity (investment), and *r* is the long-term average expected stock return. They then scale the *MM* equation by the book value of equity (B_t) at time t, and arrives at the equation

$$\frac{M_t}{B_t} = \sum_{\tau=1}^{\infty} \frac{E(Y_{t+\tau} - dB_{t+\tau})/(1+r)^{\tau}}{B_t}$$

The first implication of this relation is that a lower value of M (higher *B/M*) implies a higher expected return. This is the same insight that was presented in the value section. Two other implications of this relation are drawn in Fama and French (2015), and they are explained in the two next paragraphs dedicated to respectively the profitability and the investment factor. In these parts, the last displayed equation above is referred to as the MM equation. In short, their overall message is that the *B/M* ratio is a noisy proxy for expected returns because the market value (*MV*) also reacts to changes in forecasts of earnings and investment.

From the *MM* equation it can be seen that by holding everything else constant, higher expected future earnings imply a higher expected return. The profitability factor, as stated by Fama and French (2015), has been shown to cause problems for the three-factor model. An example of this is Novy-Marx (2013) who shows that profitability measured as gross profits-to-assets has approximately the same predictive power of cross-sectional equity returns as the book-to-market ratio. He observes that profitability is negatively correlated to value and suggests that profitability captures a different dimension of value investing than the *B/M* ratio. Specifically, whereas a value strategy finances inexpensive assets by selling expensive assets, a profitability strategy finances productive assets by selling unproductive assets.

The economic reasoning behind the investment factor also comes from the *MM* equation: holding everything else constant, higher expected growth in book equity (investment) implies a lower expected return. In their fundamental paper on capital investment behavior and stock market reactions, Titman, Wei and Xie (2004) find a negative relationship between increased investment activities and benchmark-adjusted returns[35]. Their fundamental conclusions are that investors dislike the "empire building attitudes" some firms/managers have and punish this by pushing the price down. Titman et al. (2004) further claim that the positive past returns and profitability of a company do not protect the company from experiencing a negative investment-return relation. This negative relation is explained neither by the underlying characteristics of the firms nor by the risks.

These changes in the investment behavior of firms provide information to the market, which could have positive as well as negative signals in theory. A positive signal would be that firms that invest more are related to better investment opportunities in general, whereas the fact that more investment is sometimes linked to overinvesting managers is a negative signal. According to Titman et al. (2004, p. 699), "investors tend to underestimate the importance of unfavorable information about managerial intention." They underline this by using two arguments, stating that in the 80´s the investment-return relation appeared to be reverse or at least not significantly negative[36], as well as

[35] Specifically, benchmark underperformance over the five years following the investment.
[36] Period with omnipresent hostile mergers and therefore more disciplined management.

the strong negative investment-return relations observed for firms with high free cash flows and lower debt ratios.

Fama and French (2006) also investigated the effects expected profitability and investment have on stock returns previously, but then they failed to find the links implied by the *MM* equation. Aharoni et al. (2013) show that expected investment and returns are negatively related, as implied by the *MM* equation. They show that this relation holds as long as the variables are measured at the firm level, and not at the per share level as Fama and French (2006) measure the variables. As a final note on the profitability and investment factor, it should be mentioned that the five-factor model including these two factors is fairly new in this framework.

That concludes the introduction of the individual factors. The *B/M* factor has delivered high returns in different markets over a long period and is well documented, while there is more doubt about the existence of the *size* premium due to the disappearance after its initial discovery. The momentum factor has a shorter academic history than the *size* and the *B/M* factors but has also delivered solid returns across markets and asset classes. Profitability and investment have recently been added to the five-factor model of Fama and French (2015). Motivated by the implications of the *MM* equation, it will be interesting to investigate these effects in the Chinese equity market. The thesis now moves on to some of the critique that has been presented towards the reviewed factor models.

3.4 Critique of FF

Fama and French faced various criticism for the radical interpretation of the results in their initial three-factor models (1992, 1993). There they argue that the three-factor model is a perfect implementation of the theoretical models like the Intertemporal Capital Asset Pricing Model (Merton, 1973) or Ross´s Arbitrage Pricing Theory (1976).

Among others Kothari, Shanken and Sloan (1995) argue that huge parts of the factor premia could be caused by a survivorship bias or selection bias in the underlying data sample. Following their reasoning, the relation between returns and *B/M* is unlikely to be as high and as consistent as pointed out by Fama and French (1992, 1993). Kothari et al. (1995) argue further that the relation between *B/M* and return is at best weak. Their conclusion comes from testing using a different composition of market beta, observation frequency as well as a different data source.

MacKinlay (1995) suggests that the found premia are a result of “data snooping.” He shows that an ex-ante prediction of CAPM deviations due to common risk factors might be hard in an empirical set-up.

Non-risk based deviation sources should be taken into consideration, as they are easier to detect. For him, factor models are not able to explain the entire deviation picture. To assess ex-ante CAPM deviations due to risk factors, MacKinlay (1995) bases his analysis on mean-variance efficiency theory. By doing so, he can control for the increased variance that typically accompanies the deviation of expected returns.

The mean-variance analysis is chaperoned by zero-alpha F-tests introduced by Gibbons, Ross and Shanken (*GRS*) (1989) and MacKinlay (1987). Further, MacKinlay (1987) argues that only in the case that found premia are caused by underlying risk, the Sharpe ratio (the typical risk measure in portfolio theory) is an appropriate measure as it is upward bounded in this case. If there is a different underlying cause for the premia than risk, the Sharpe ratio will not be an appropriate measure since it is not upward bounded. MacKinlay is among the first to implement the F-distribution based *GRS* test and thereby provides a useful test for robustness and comparative statistics among different risk-based models.

Fama and French (1996) use the *GRS* test to assess the precision of their models. In their paper, they find that low *E/P* (Earnings-to-price) and *C/P* (cash-flow-to-price) have the same connection to financial distress and expected returns as low *B/M* ratios. For one- and two-dimensional sorts the *GRS* test fails to reject the H0-hypothesis of intersects being zero. Due to this, Fama and French (1996) conclude that the three-factor model covers almost all the variation in the expected returns. They verify that the additional methods and ratios used to form dependent variable portfolios are not able to uncover additional risk dimensions and expected returns other than those required to explain the returns on the independent variable portfolios in the original Fama and French set-up.

Titman and Daniel (1997) provide a critical review of the overall conclusion Fama and French draw in their initial three-factor model paper (1993). According to Titman and Daniel (1997), asset pricing should focus more on the underlying characteristics of the companies than on the covariance structure when trying to explain excess returns. They investigate whether returns to *B/M* and *size* can be explained by the factor loadings of stocks. Despite a strong covarying behavior of stock returns of companies with similar *B/M* ratios Titman and Daniel (1997) cannot find any evidence for this being caused by increased risk patterns associated with financial distress scenarios. They rather suspect this covariance to stem from the similar underlying business environment (industries, countries, regions). When controlling for certain underlying firm characteristics, there was no explanatory power left to the factor loadings. In their analysis, Titman et al. (1997) show that market betas are close to one and therefore as already mentioned by Fama and French (1993) have very low to no power in terms of explaining different cross-sectional equity returns. Factors fail to explain returns above the threshold unto where factors are pure proxies for the underlying firm characteristics. If risk is not the main component explaining returns, but returns are rather explained by a company's fundamental characteristics, the betas of the estimated models/ factors might contribute very little to the attempted explanation of the return variation.

Davis, Fama and French (2000) take the critique of Titman et al. (1997) into account and test with a similar set-up for an extended period[37]. They claim that they cannot find convincing evidence for the characteristics model to explain the expected returns in a better way than the risk model. The three-factor risk model does a better job than the characteristics model. On the other hand, they also admit that a joint test performed to test the overall fit of the model (following Gibbons et al., 1989) rejects the three-factor model. They consider this a normal drawback arising from working with simplified models. In general, expected returns should reward for risk loadings independently from the underlying *B/M* characteristics (Davis et al., 2000). Titman et al. (1997) are focusing on the relative distress of companies, and how it drives stock returns and can be measured using the *B/M* ratio as a proxy (high *B/M* is associated with financial distress). Returns do not rely on the risk exposure. On the other hand, Davis et al. (2000) are focusing on the underlying risks itself, picturing how the expected returns reward investors for taking on risk.

In a follow-up paper Wei, Daniel and Titman (2002) test their "characteristics-model" out of sample on the Japanese market[38]. In the theoretical set-up of their model, they introduce an equation[39] where the expected future returns are a linear equation of the intercept (or alpha) and the characteristics component as well as the factor loading component. In the characteristics-model world of Wei et al. (2002), the last component (the factor loading) is equal to zero, and the expected future returns appear to be described by a simple linear function of the characteristics factor. In the "risk-model" of Fama and French (1992, 1993), the reverse is true. The characteristic component is zero, and the expected future returns are preliminarily described by the risk-premia component. Wei, Daniel and Titman (2002) mention that a case in which both components are non-zero would be plausible. Furthermore,

[37] Davis et al. (2000) use the period from 1929 to 1997 to perform an out of sample test of Titman et al. (1997) characteristics model. To ensure the out of sample characteristics they skip the original observation period (1975-1997)

[38] They choose the Japanese market due to data availability over the same time horizon (1975 – 1997) as used in their first paper, as well as the strong Japanese value premium (performance of the Japanese HML factor is almost triple the size compared to the US market)

[39] $E_{t-1}[R_{i,t}] = \alpha + \delta\theta_{i,t-1} + \lambda ß_{i,t-1}$, this formula determines the expected returns used in factor models and is part of the overall factor model equation: $R_{i,t} = E[R_{i,t}] + ß_{i,t-1}f_t + \epsilon_{i,t}$

they point out that in many cases it might be very hard to distinguish between the "risk-model[40]" and the "characteristics-model" due to high multicollinearity. According to them, this will be worsened by the way Fama and French select and set-up their portfolios, as they sort all companies with similar *B/M* ratios and sizes into one portfolio.

To be able to distinguish between the characteristics and the risk effect Wei et al. (2002) form portfolios where they combine high *B/M* ratios (which for them are associated with a high characteristic θ) with high and low factor loadings (ßs) on the HML factor and low *B/M* with high and low factor loadings. If the characteristic model holds, these high *B/M* portfolios should deliver high returns, whereas they would deliver zero returns on average if the risk-factor model of Fama and French holds. Wei et al. (2002) are able to reject the factor model for the US and the Japanese market (for both markets for the period 1975 – 1997) but are unable to reject the characteristics model. In the US-case Titman and Daniel (2002) are able to reject the factor model for all cases of characteristic-balanced portfolios (loading on HML, SMB and market), whereas in the Japanese case they can only reject for the characteristic-balanced portfolios loading on HML. In general, it seems to be a bit arbitrary how test periods and markets are selected by different studies. As also mentioned by Wei et al. (2002) in their conclusion, for some periods and markets it might be almost impossible to distinguish between the fundamentals of the models (i.e., underlying characteristics and underlying risk factors/factor loadings).

That concludes the general part of the literature review. Fama and French (1993) have influenced much of the research of the last decades. A wide range of research has been conducted based on the techniques Fama and French initially developed and the conclusions they drew based on their findings, but as seen, the paper has also been subject to much critique. Nevertheless, due to the influence, the paper has had on factor research, the techniques later described in the methodology and applied in the analysis are to a large extent based on the techniques developed by the initial Fama and French (1993) and the later Fama and French (2015) papers. To further investigate if the general techniques should be modified in any way, factor research on the Chinese equity market will be reviewed in the next part.

[40] Also often called the factor model

3.5 China

The combination of a rapidly developing equity market and the increasing economic importance of China in the world economy has led to increasing interest in the Chinese equity market. However, factor model research in China has largely been limited by the relatively short history of the stock market, as well as availability and quality of data[41]. Much of the existing literature utilizes different methodology, data sources, and data handling, and as a consequence of that achieves conflicting results. Therefore, this section aims to review some of the most relevant and influential papers on the topic to investigate how market peculiarities are accommodated for and highlight the market features broadly agreed upon as well as the features the literature is conflicted about. An additional aim is to investigate how the literature accommodates special features of the Chinese market. To achieve this, the section starts by describing the earliest factor research on the Chinese market and follows the development of the research up until the current point in time. The main weight is nevertheless put on the most recent research since it is the most relevant.

Drew et al. (2003) were among the first to explore the explanatory power of the *Size* and *B/M* factors in the Chinese market. They set out to investigate the multi-factor approach for equities listed on the Shanghai Stock Exchange in the period from 1993 to 2000. That is, they include A-shares (tradable and non-tradable) and B-shares for all listings on the exchange. Furthermore, already here they note concerns about the definition of model variables due to the complicated share structure in China. These concerns are related to the valuation of non-tradable shares and B-shares and through that the total market value of a firm.

In line with research on the US market, they find that the market factor is not sufficient to explain the cross-sectional variations in average stock returns on Shanghai Stock Exchange and that the size variable is highly useful as an additional explanatory variable. In conflict with the existing literature on different markets, they find that *B/M* does not appear to be useful. In fact, they find that growth firms generate a superior return when compared to value firms, contrasting the research on the US market where value firms have been shown to generate superior returns. Drew et al. (2003) further offer several possible explanations for the growth premium they document, among others overexploitation of the value effect, irrational investor behavior, or special features of the Chinese market. They specify these special features as the extent of government intervention, investor composition and market structure in terms of the many different share classes and non-floating shares. Without looking further

[41] See for example Hu et al. (2018)

into this, they propose that the explanation might lie somewhere in the cross-sectional differences in non-traded and institutional holdings.

Gan et al. (2013) provide supporting evidence to the findings of Drew et al. (2003) by finding significant size- and negative *B/M* premia between 1996 and 2005. It is worth noting that the results displayed by Gan et al. (2013) differ significantly from both Drew et al. (2003) and Fama and French (1993) in one aspect: their models deliver much lower adjusted R-squared values than the two latter for the three-factor model.

In the following year, Wang and Xu (2004) investigate the importance of the non-traded fraction of a company's total shares. Drew et al. (2003) also mention this as a possible reason for the apparent lack of information in the Chinese *B/M* factor. First, Wang and Xu (2004) find that the size factor is useful for explaining differences in time variations for non-financial A-shares in the period from 1996 to 2002, while the *B/M* factor is insignificant. As previously described in part 3.3 on individual factors, the *B/M* factor normally proxies for fundamental value in a firm. Wang and Xu (2004) therefore argue that its explanatory power weakens when either the book value of a firm is difficult to assess, or investors do not pay much attention to fundamentals. Hence, they argue that the weak explanatory power of the *B/M* variable makes sense in a Chinese investment environment characterized by dubious accounting standards, as well as speculative and irrational investor behavior. Second, they also replace the *B/M* variable with a float ratio variable. Following Gomper et al. (2003), which suggests that better corporate governance signals better performance in the long run, they suggest that the percentage of floating shares is a good proxy for corporate governance. Specifically, stocks with higher float ratios could signal higher future cash flows, and through that also higher returns. They further show the float ratio to be a significant factor in the explanation of return variations across stocks and therefore suggest that the *B/M* factor in the traditional Fama and French three-factor model should be replaced by a float ratio variable to accommodate for the unique Chinese equity structure.

The papers reviewed so far on the Chinese market have only covered periods reaching approximately up to the share-structure reform in 2005. Moving on to papers reaching beyond the share-structure reform, Xu and Zhang (2014) represent a bridge between the literature covering the period before and after 2005 by investigating the period from July 1996 to June 2013, as well as the two sub-periods from 1996 to 2004 and 2007 to 2013. However, their main period of investigation is from 1996 to December 2004. For that period they estimate the three-factor model in the Chinese stock market and find both size and value premia. Previous literature has been inconsistent on issues related to the construction of the three variables, and Xu and Zhang (2014) therefore investigate several of these issues. Their findings lead to several useful methodical suggestions:

1. The market portfolio should be constructed only by including tradable shares;
2. Dividing firms into size groups should be done based on total market value;
3. The Small and Medium Enterprise Board (SME) and the Growth Enterprise Board (GEB) stocks should be included to determine portfolio breakpoints;
4. The book value per share-to-price per share ratio should be used instead of the *B/M* ratio.

Furthermore, their investigation of risk premia over different periods shows that the estimated risk premia depend heavily upon the period of investigation. As described, they split their longest period into two additional sub-periods. While the early sub-period and the entire period look very similar in terms of means, standard deviations and Sharpe ratios for the factor risk premia, the reported values for the later period are very different. For example, whereas the whole period and first sub-period results yield sizeable positive risk premia for all the three factors, the later sub-period yield negative risk premia for both the market and the *B/M* factor. This illustrates how dependent on the construction of the variables estimated risk premia are in the Chinese market.

Cheung et al. (2015) investigate the Chinese A-share market as represented by the MSCI China A share index in the period from 31st of December 2001 to 31st of December 2013. Specifically, they investigate the existence of a significant premium to five well-known factors from international research: value, size, momentum, dividend yield and volatility. They further check if these factors contribute to explaining the variance of returns to a larger extent than the CAPM. Their main findings are significant (5% level) value and dividend yield premia, as well as positive but insignificant (5% level) premia on CAPM beta, size, momentum, and volatility. In addition, they find that the Fama and French three-factor model with market-, value- and size factors significantly increases the explanatory power of the variability in Chinese equity returns compared to the CAPM. The other three factors investigated do not add anything extra from this perspective. The lack of a significant market premium is inconsistent with much of the Chinese factor research, and Cheung et al. (2015) attribute the lack of significance in their sample to the high volatility of the market.

The lack of a significant size premium is consistent with the Fama and French (2012) findings for developed markets but inconsistent with most previous research on the Chinese market. Cheung et al. (2015) attribute the insignificant size premium to the high volatility of small-sized securities. Furthermore, they state that the difference in their findings from other papers could be due to different methodologies and time periods. The significant value premium is consistent with the findings of Fama and French (2012), while there is mixed evidence on the value premium in the Chinese market.

Conflicting with the findings of Fama and French (2012), they find no momentum premium in the Chinese market. Puzzling enough, given the short-term nature and high turnover of the Chinese market, research on the Chinese market has mostly been unsuccessful in detecting a momentum effect. As stated in section 2.2, the average annual turnover is reported to be approximately 500%, which corresponds to an average holding period of around two months. Therefore, Cheung et al. (2015) suggest that the momentum factor is constructed based on a too long period in the Chinese market. Kang et al. (2002) find significant evidence for a momentum premium in the Chinese market using weekly returns for 12- and 26-week factors, but in general, there is a lack of evidence in favor of a momentum factor in China. Furthermore, Cheung et al. (2015) state that it seems highly sensitive to the time horizon.

Xie and Qu (2016) use monthly data between 2005 and 2012 for listings on the Shanghai Stock Exchange (*SSE*) and find that the Fama and French three-factor model to a large extent is appropriate for explaining the stock returns in their data set. Furthermore, they find indications of both *size*- and a value-premia. In terms of construction of the independent variables, Xie and Qu (2016) use the SSE A-share Index as the market portfolio return, the floating shares market value for the *size* factor, and the traditional *B/M* ratio as balance sheet book value-to-total market value. That is, they differ from the recommendations of Xu and Zhang (2014) on the two latter variables since Xu and Zhang (2014) recommend to use total market value for the *size* factor and the book value per share-to-price per share for the value factor. Xie and Qu (2016) argue that using the market value in circulation is correct when sorting the size portfolios due to the different price behavior of tradable and non-tradable shares.

Chen et al. (2015) investigate the Chinese A-share market from 1997 to 2013, and later they update their results in Hu et al. (2018) by extending the period to 1995-2016. They find a significant *size* premium, but no significant equity and value premia in both papers. However, both the equity and value premia are positive, with a sizeable equity premium and a small value premium. They perform several robustness tests, and their results are overall robust against changes in variable construction and weighting regimes. However, when they change the period of investigation, the value factor becomes significant. Because of that, they conclude that the differing results in the literature concerning the value factor mainly are due to different periods of investigation.

Hu et al. (2018) follow the methodology of Fama and French (1992, 1993) closely. First, they perform a time series regression following Jensen et al. (1972) and Fama and French (1993) to check if the market-, size- and *B/M* factor capture common variations in stock returns. Then they perform a cross-sectional analysis inspired by Fama and MacBeth (1973) to estimate the factor premia. It is worth noting that Hu et al. (2018) differ from the conventional methodology when estimating the factor premia, as they estimate variance adjusted average returns instead of value-weighted or equally weighted returns. Specifically, they weight the portfolio returns by the inverse of the variance estimated from the portfolio's daily returns over a 3-month rolling window. This is done since the three factors do not seem to have the same dispersion across time. By variance-weighting the returns, they put more weight on quiet periods (high precision) while putting less weight on periods with high market volatility (low precision). The *size* factor is constructed by sorting firms by their floating market capitalization, while they follow the advice of Xu and Zhang (2014) by constructing the *B/M* factor as book value per share divided by the price of one A-share.

Lin (2017) closely follows the methodology of Fama and French (2015) to test the five-factor model in China. He investigates all firms listed in the Chinese A-share market from 1997 to 2015, and find that the five-factor model provides a better description of Chinese equity returns than the three-factor model. Furthermore, he finds the investment factor to be redundant. In contrast to this finding, Belimam, Tan and Lakhnati (2018) find the three-factor model superior to the five-factor model in China. However, they only investigate Shanghai A-shares over the period from 2011 to 2016.

To sum up, research on the period before the share-structure reform in 2005 seems to agree on the existence of a size premium and that a three-factor model explains more of the observed stock returns in China than the CAPM model. The value premium is on the other hand non-existing, and the existing research indicates that if there is a premium related to the *B/M* ratio, it is rather a negative value premium. Wang and Xu (2004) also successfully replace the book-to-market ratio with a governance proxy, namely the float ratio. This model shows good explanatory power in the Chinese market before 2005.

For periods reaching beyond the share-structure reform of 2005, there is mixed evidence. The three-factor model explains much of the variation in this market, and it is able to explain more of the variation in Chinese stock returns than the CAPM model. Cheung et al. (2015) find a significant value premium but no significant equity and size premia in the period from 2002 to 2014, Xie and Qu (2016) find indications of the presence of all three premia for the period from 2005 to 2012, and Hu et al. (2018) find a significant and robust size premium but no robust and significant equity and value premia from 1995 to 2016. Besides the investigated premia in this thesis, Cheung et al. (2015) also find significant volatility and dividend yield premia. Two papers testing the five-factor model in China have also been reviewed, where one concludes that the five-factor model is superior in the Chinese market and the other argue the superiority of the three-factor model. However, the two papers investigate two very different time periods, and their constituent's lists are also different. It is therefore difficult to compare them directly.

This concludes the entire literature review. As mentioned previously, the methodology of this thesis will be heavily based on Fama and French (1993) and Fama and French (2015). In addition, one of the stated sub-questions in section 1.2 addresses whether the market structure of the Chinese market implicates that general methods should be altered. After outlining the Chinese equity market and the Chinese factor literature, it is a good time to address this question. Many peculiarities of the Chinese A-share market are outlined in section 2.2, but most of them are something to bear in mind while interpreting the observed effects rather than something that will be accommodated for in the methodical setup. Some of them are also difficult to accommodate for due to limited and differing accessibility ranges of data.

Something that is necessary to accommodate for is the tradable and non-tradable share structure. Several different methods to form the factor portfolios are applied in the literature, especially on the use of tradable or total market cap for the formation of the size factor and the use of the traditional *B/M* measure or the China-specific book value per share divided by price per A-share. However, the thorough robustness checks performed by Hu et al. (2018) might indicate that these differences are of less importance and that it is rather the differences in the period of investigation that cause the wide dispersion of results. Nevertheless, this thesis also takes a stance on how to form the factors.

In short, this thesis follows the recommendations of Xu and Zhang (2014), since they specifically look at the methodical issues related to the share structure. That is, the market portfolio is constructed by including tradable shares, size is constructed based on total market value, book value per share-to-price per share is used instead of the standard *B/M* variable, and the SME and GEB stocks from Shenzhen stock exchange are included. All of these choices except the inclusion of SME and GEB stocks will be further elaborated on in the methodology and data sections where the construction of variables are more closely described. The thesis now moves on to the specifics of the methodology in the next section.

Section 4 – Methodology

The methodology of this master thesis is in general closely linked to the methodology applied by Fama and French (1993, 2012 and 2015) and Carhart (1997). It assumes that the reader has a basic knowledge of modern statistical approaches and frameworks.

It is structured as follows. First, the section starts with a broader and more general look at research and methodology. Then it proceeds with a general description of regression techniques including a description of the OLS framework, statistical tests and time series characteristics. That is followed by a short introduction to the theoretical background and technicalities of the models used in the analysis. The fourth subsection describes the construction of the variables used for portfolio and factor creation. These variables are the basis for the portfolio sorting and creation of the dependent variable portfolios in the following fifth subsection. The sixth subsection concerns itself with the creation of the factors used as the independent variables in the models before the seventh section describes which models are tested and how dependent and independent variables are used in the models. At last, the overall measure of fit used for evaluation of the models in this thesis is described.

4.1 Philosophy of research

This subsection begins with the philosophy of research for social sciences and how this philosophy relates to defining the overall aspects of a research project. It then narrows in by defining this research project along four different dimensions, and further by describing the specific approach taken here.

At the overall level, there are three different approaches to social science. These three approaches are important for the choice of the overall methodology and represent different paradigms of research.

The concept of paradigms was first popularized by Thomas Kuhn and has to be interpreted as an entire system of thinking. That is, it roots in the same underlying assumptions, research methods, and focus of research. Kuhn argues that science cannot be compared across paradigms. Another influential person in the field of scientific philosophy, Karl Popper, on the other hand, argues that science can be compared across paradigms, even though he admits that it may be difficult.

In the following, the three above approaches to social science are explained. The first approach to social science research is *positivism*, which is the traditional natural science approach. This is the most common perception of science in general, and it aims to unveil the facts of an area. These facts form the basis for the prediction of resulting outcomes.

The second approach is *interpretive social science*. This line of methodology focuses on understanding social interaction as a part of the context it appears in and describing it as such. Therefore, it represents a natural opposition to the harder universal laws positivism investigates. As a concrete example of the two, positivists at the most extreme would investigate the behavior of people by measuring what they believe to be important and quantifying it regardless of the context it is measured in. On the other hand, interpretative researchers would focus on understanding the full context behavior occurs in. The classic example of how to achieve this is a scientist living amongst the subjects of research to understand their thoughts and feelings.
Critical social science is the third approach to social research. It criticizes both the two approaches above, the positivist approach for too bombastically looking for absolute and universal truths, and the interpretative approach for being too narrow and focused on people's perceptions instead of actual conditions. Therefore, it places itself somewhere in between the two other approaches, with a more pragmatic approach that depends on the needs of a specific situation. A distinct feature about the critical approach is that it specifically focuses on changing society for the better by breaking common illusions. All three approaches further have several under categories, which will not be explained in further depth.

This thesis mainly takes a positivistic approach, even though the research performed also contains some aspects of the critical social science approach.

Neuman (2014) highlights four different dimensions of research that should be defined. The sum of these four dimensions defines what kind of philosophical direction a research project takes.
The first dimension is the purpose of the study; is it exploratory, descriptive or explanatory. That means defining if the purpose of the research is exploring a new field, describing an already explored field in greater detail, or explaining the reason to something that is already explored and detected.
The second dimension to define is what the study is intended for. Neuman (2014) defines two possible types of research: basic research types intended for a scientific community, or applied research types intended for the general public, participants, generalist practitioners or narrow practitioners.
The third dimension to define is the time aspect of the study. That is, studies can investigate across subjects at the same point in time (cross-sectional), the same subject over time (time-series), or combinations of the two (panel data, cohort analysis, case study). Different sides of the three first dimensions are commonly used for all of the three different philosophical directions previously described, and there is not one or another approach that is uniquely tied to a philosophical direction.

The fourth and last dimension defined by Neuman (2014) is the data collection technique. Closely corresponding to the positivist and interpretive social science philosophical views of science described above, research is typically either quantitative or qualitative. Quantitative research focuses on measuring objective facts that are seen as independent of context, while qualitative research focuses on situationally constrained social reality. Quantitative research therefore typically analyze many observations applying statistical analysis, and qualitative research concentrates on understanding fewer subjects in greater detail.

Due to the nature of stock market data and the accessibility to data in China, the methodology and research philosophy applied is naturally guided towards a positivistic and a quantitative, mostly time series based, exploratory, scientific community directed research approach.

This thesis moves in the territory between exploratory and descriptive research. It is probably more descriptive than exploratory, as factor models in general are widely described. However, the topic is not extensively explored in the Chinese A-share market. The thesis sets out to provide detailed descriptions of the market at the same time as focusing questions for future research. The purpose of this study is also somewhere between the two camps of basic research and applied research. It tilts more towards basic research, as it tries to explain how the Chinese market functions. At the same time, it also has an applied purpose as it could be of particular value for generalist practitioners and specialist practitioners. Furthermore, it deals with both cross-sectional and time-series data and is to a large extent quantitative.

This thesis is of a quantitative nature, and the variables investigated is what Agresti and Finlay (2014) categorize as quantitative response variables. Following their guide[42], descriptive methods, as well as univariate regression, multivariate regression, and correlation methods are applied. These methods will be thoroughly described in the rest of the methodology, as the whole process closely built upon Fama and French (1993) is outlined.

4.2 Data treatment

The description of methods applied begins with the specific methodology for data handling since this thesis deals with extensive amounts of data. To perform quantitative research, this thesis collects a large amount of data on different variables of interest. The raw data needs to be treated in a way that

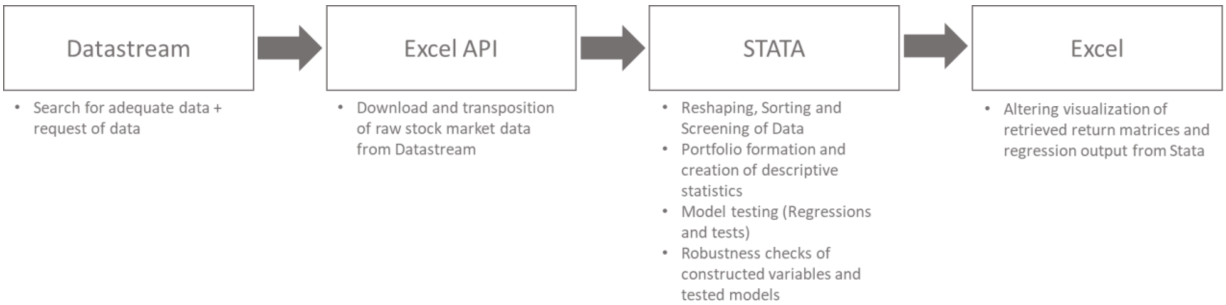
[42] Agresti and Finlay (2014, page 551)

makes it possible to analyze it in an adequate academic manner. Cleaning, restructuring, and reshaping of the data are one of the essential issues that need to be performed before performing quantitative analysis.

Programs used to process data

For data processing, data cleaning, portfolio formation, statistical tests, and other data and statistic related issues, Stata® (version 14.0) is used. An extensive code is written and organized in different Do-Files and Log-Files. For reasons of readability, no code will be displayed in the upcoming sections. The entire code is documented in a separate file attached to the thesis. Writing the code as flexible as possible is a focus of this thesis, as this set-up allows for simple changes of underlying variables, reassessment, and extensive robustness checks. Furthermore, data treatment and portfolio formation in a non-code based program as Microsoft® Excel is not an appropriate program to address the underlying issues of this thesis, as it does not allow for appropriate documentation of work stages and the desired flexibility. Dealing with a large amount of data could lead to several problems in the scope of scientific work. For reasons of reproducibility, testability, maintainability, and accuracy as well as traceability of errors, the code is toroughly documented using comments. In this thesis, Microsoft® Excel is mainly used as an API to access Datastream data and format the layout of the displayed tables[43].

Figure 1 - Data handling process



4.3. Statistical framework

Regression Techniques[44]

As the analytical section of this thesis is mainly based on an empirical framework, descriptive statistics, performance of regressions and robustness checks of the underlying models need to be theoretically

[43] Unfortunately Stata® does not provide much options to customize the layout of produced outputs
[44] This section mainly refers to: "Introductory to Econometrics – A Modern approach 4th Edition" by Jeffery Wooldridge

motivated and justified. This is done by standard approaches and methods used in a general statistical framework and complemented by more specific methods and tests. In this empirical framework, quantitative techniques are used to estimate potential causal relations between dependent and independent variables.

Descriptive statistics

The first step in the scientific analysis is to obtain an overview of the underlying data used in the analysis. Many of the variables of interest can be displayed using simple graphical illustrations, which display initial trends in the data. Simple averages, correlations, standard deviations, and t-statistics further help to formalize these trends.

The OLS Framework

This thesis focuses on the evolution of asset pricing techniques over the last 30 to 40 years and its empirical implementation on the Chinese Stock market. Asset pricing models are mainly based on the relation between past returns and underlying characteristics of the market where one tries to estimate the causal effect a specific independent variable/ factor has on the dependent variable. This relation is typically estimated by using standard Ordinary Least Squared (*OLS*) regressions. Both univariate and multivariate OLS regressions fit a model based on one or several dependent variables into a cloud of observed values, by minimizing the sum of squared residuals. That is, the OLS regression minimizes the squared deviations from the estimated line for each of the predicted variables.

$$y_i = \alpha_i + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

For the OLS regression to deliver unbiased estimates, several conditions must be fulfilled. These conditions are mostly referred to as Gauss-Markov assumptions. In the simple regression case (univariate regression), the assumptions include:

1.) Linearity in parameters: a regression is determined by the linear relationship $y = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \cdots + \beta_k x_k + \mathrm{u}$,

2.) Random sampling

3.) Sample variation in the explanatory variable/ no perfect collinearity between independent variables as well as no exact linear relationships

4.) Zero conditional mean: $E(u|x_1, x_2, \dots\ x_k) = 0$ The error term *u* has an expected value of zero

5.) Homoscedasticity: The error-term u has the same variance given any values of the explanatory variable $Var(u|x_1, \dots\ x_k) = \sigma^2$

6.) If the sample error-term *u* is additional to the Gauss-Markov assumptions, also independent from the explanatory Variables and normally distributed with zero mean and variance σ^2 a Classical Linear Model (CLM) is estimated

The first four assumptions are typically not an issue, especially when using an appropriate statistical program like *Stata*. *Stata* estimates the asymptotical efficiency of the OLS assumptions and corrects if the dependent variable is not approximately normally distributed. Under the condition that the five first Gauss-Markov assumptions hold, an OLS estimator is called Best Linear Unbiased Estimator (BLUE).

Standard Hypothesis tests

In a statistical framework, certain standard tests are used to assess and compare the influence of different dependent and independent variables in single and joint tests.

T-tests

In general, a t-test tests whether an estimated value or mean is different from a specific value.

$$t_i = \frac{\hat{\beta} - \beta_0}{SE[\hat{\beta}]}$$

, where $\hat{\beta}$ is the estimated value, β_0 is the value to test against[45] and $SE[\hat{\beta}]$ is the standard error of $\hat{\beta}$

This thesis mainly uses t-tests to evaluate whether means of premia and other variables are different from zero and whether the coefficient estimates of the regression models are different from zero.

If all the six conditions above hold, the t-distribution applies for the standard estimators and unbiased t-statistics can be estimated. This is an important property as the t-distribution provides the base for

[45] In this thesis mainly zero

H_0-hypothesis testing. If the conditions above do not hold the t-statistics are not unbiased and need to be corrected. To avoid issues arising from biased t-statistic estimations, this thesis uses robust standard errors whenever t-statistics are calculated.

F-tests

This thesis does not apply F-tests directly, but the properties of the F-distribution are important as they help to build the framework for the later used Gibbons-Ross-Shanken (*GRS*) test. F-tests are one-sided joint tests that are based on the F-distributions.

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{(SSR_{ur})/(n-k-1)}$$

, where n-k-1 is the degree of freedoms of the nominator and q is the degrees of freedom of the denominator and SSR is the sum of square residuals

The F-statistic typically tests the joint H_0-hypothesis that all coefficients (*ß*s) are zero and therefore provide no explanatory power. Therefore, it is an indicator of the overall significance of the regression. That is, the H_0 in this framework is given by

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

The F-statistic can be estimated from the *R-squared* of the regression and is essentially the relative difference between a restricted and an unrestricted model, corrected for degrees of freedom.

Common issues with OLS-Regression

After introducing the standard conditions and tests within the framework of regression techniques some of the issues commonly appearing when dealing with OLS regressions are discussed more detailed.

Homoscedasticity

The presence of heteroscedasticity violates the fifth Gauss-Markov assumption. This means that the estimators are no longer *BLUE,* and the biased standard errors tend to reject the H_0 more easily. This issue can be fixed by introducing heteroscedasticity robust standard errors. All regression estimations in this thesis use heteroscedasticity robust standard errors, although heteroscedasticity does not appear to be a severe issue in the data sample. Using robust standard errors although there might be no homoscedasticity does not harm the estimation results.

Multicollinearity

In general, multicollinearity is defined as high correlation between two or more (independent) variables. This high correlation might lead to higher standard errors as well as lower explanatory power of the single betas among the highly correlated variables. Typically, that makes it more difficult to extrapolate the partial effect of each of the estimated coefficients.

Omitted Variable Bias

Omitted variable bias stems from the exclusion of essential independent variables from the model, which leads to an underspecified model. Hence, the model does not reflect the true underlying relation between dependent and independent variables in the data anymore, as omitted variable bias leads to biased coefficients.

Both multicollinearity and omitted variable bias can be an issue in the scope of a factor model analysis. If one only seeks to minimize alpha and the error terms, this might not harm the analysis severely. Omitted variable bias is rather assumed to appear in lower degree factor models. This will bias the coefficients and leave the model with a higher amount of unexplained variation. In general, there is a trade-off between multicollinearity and omitted variable bias.

Time series Set-up

As the underlying analysis in this thesis mainly is based on time series regressions rather than cross-sectional analysis, the following section focuses on the properties of time series analysis. The dependent variables to be explained by the models used in this thesis are returns. First, some of the most common issues appearing when dealing with time series are highlighted in this section.

Stationarity

A time series process can be either stationary or non-stationary. Whether a process is stationary or non-stationary has certain implications for the process and the estimates based on this process. A stationary process has the same probability distribution over time. Non-stationarity makes estimation methods and forecasting imprecise and often lead to “exponential trending” behavior of the underlying variable over time. In general, most prices tend to follow non-stationary processes, whereas their first differences are stationary. To eliminate nonstationarity, it is required to form differences in logs[46]. Most of the asset pricing literature uses simple returns. They are assumed to be at least weakly stationary and therefore provide unbiased estimates of the mean and variance. Simple returns show noisy behavior as they tend to vary around a fixed line. For stock returns, this is typically the time-axis, at least on a long-term perspective.

Trends

In general, two types of trends exist: stochastic trends and deterministic trends. Trends are long-term behavior/ movements of a variable over time. A deterministic trend is a non-random function of time (i.e., a straight line over time). A stochastic trend is random and varies over time. It may have prolonged periods of increases or decreases, and stochastic trends are therefore typically used in finance and econometrics to model time series and perform forecasting. The random walk is probably the most famous stochastic trend, and especially random walks with drift are present in stock prices. However, as previously stated, this analysis uses return data. As explained in the previous section, returns are assumed to be weakly stationary, and this issue should therefore not be present. Returns in the underlying case are most similar to a white noise process.

Autocorrelation

A series that is correlated with its past values is said to be auto-correlated or serially correlated. Regressions estimated under autocorrelation still produce valid OLS estimators but standard errors might be biased. There are several ways to test for auto-correlation, including estimations of (partial) autocorrelation functions, testing the F/t-distribution and R-squared of regressions performed by including several lags of the dependent variable, Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC).

Fama and French (1993) show that autocorrelation within the US sample was only present to a minor degree. Drew et al. (2003) found the same for the Chinese market, leading to the conclusion that lagged independent variables should not be part of the factor-regressions. While testing for time series specificities, this thesis finds no evidence for severe differences that would involve deviation from the aforementioned cross-sectional methods and models.

[46] $diff_P = \ln(P_t) - \ln(P_{t-1})$

4.4 Specific models used[47]

The topics covered so far in the methodology has touched upon the general issues related to statistical analysis. This is the theoretical foundation for a time series based analysis. The following section concerns itself with the theoretical background, development, and implications of factor models and factor-based investment. These are the building blocks for the models estimated in this thesis.

The goal of this thesis is to unveil the return structure of the Chinese equity market by means of risk-based factor models. Risk-based factor models are deemed as an extension or substitute of the Capital Asset Pricing Model (*CAPM*). Like the *CAPM,* these models seek to explain the returns in an equity market by the exposure of a certain asset to an underlying risk factor. These risk factors or risk premia proxy for certain essential characteristic of companies. Among others, fundamental work in this field was done by Robert Merton (1973) and Stephen Ross (1976). Merton's Intertemporal Capital Asset Pricing Model and the Arbitrage Pricing Theory (*APT*) developed by Ross extend the classical *CAPM* by additional risk factors.

The classical *CAPM* seeks to explain expected returns of an asset as a cross-sectional relationship. This is illustrated by

$$E[R_i] = r_f + \beta_{iM}\left(E[R_M] - r_f\right),$$

where r_f is the risk free rate in the market (typically a treasury note or a short term government bond), β_{iM} is the exposure of asset *i* to the market (its risk exposure) and R_M is the expected return on the market.

From the equation above, the empirical time series set-up of the CAPM follows as

$$r_{t,i} - r_{t,f} = \alpha_i + \beta_i\left[r_{t,M} - r_{t,f}\right] + \varepsilon_{t,i},$$

with

$$E[\varepsilon_t] = 0$$

$$Cov[R_M, \varepsilon_t] = 0,$$

[47] In addition to the standard papers, the thesis follows the approach/ argumentation of Claus Munk (2017) and Campbell, John Y., Andrew Wen-Chuan Lo, and Archie Craig MacKinlay (2012).

where for all t {1,…,N}, r_f is the risk free rate in the market, *alpha$_i$* is the intercept of the model[48], β_{iM} is the exposure of asset *i* to the market (its systematic risk exposure), r_M is the return on the market and $\varepsilon_{t,i}$ is the idiosyncratic risk of an asset (not related to the systematic/ market risk).

Through the lens of factor models, the *CAPM* is a factor model with only one risk factor, the market risk. A multi-factor (also called K-factor model) model in an empirical/ time series set-up is given by

$$r_{t,i} - r_{t,f} = \alpha_i + \beta_{i1} F_{t,1} + \cdots + \beta_{iK} F_{t,K} + e_{t,i},$$

where *alpha* is the intercept, the βs are the exposure of an asset to risk factors, *F* represent common risk factors and *e* is the idiosyncratic risk of an asset *i*.

Ross (1976) bases his *APT* model on this underlying risk connection. He shows that in a world with no arbitrage opportunities and perfect risk diversification possibilities (enough tradable assets are available to diversify all idiosyncratic risk) the model effectively explain returns. This essentially transforms the K-factor model into the *APT* and defines it by

$$E[r_i] \approx r_f + \beta_{i,1} RP_1 + \cdots + \beta_{iK} RP_K,$$

where RP_K is the risk premium for bearing factor *K*s risk, and β is the exposure to the risk premium.

The difference between *F* and *RP* is that *RP* is a risk premium that rewards for carrying a specific risk and *F* is an underlying risk factor that commonly affects all returns.

Fama and French used the APT and Multi-factor framework for building their three-factor model. This model appends the CAPM or Single-Index-Model with two other factors. As pointed out before, Fama and French chose these factors based on previous research on underlying characteristics and return patterns. The time series set-up of the Fama and French 3-factor model is given by

$$r_{t,i} - r_{t,f} = \alpha_i + \beta_{i,m}\left(E[r_m] - r_f\right) + \beta_{i;SMB} SMB + \beta_{i,HML} HML + e_i,$$

where SMB and HML are factor mimicking portfolios, *alpha* is the outperformance of the observed returns compared to the returns predicted by the model, and the *ß's* are the factor loadings/exposures to the risk factor

The estimation is done by a time-series regression. That is, regressing the left-hand side (*LHS) excess* returns on the three risk factors. In order to run this regression, several preliminary steps are necessary, including portfolio sort and portfolio return generation. The methodology first proceeds

[48] Sometimes also called the additional return in excess of the *CAPM* risk award

with outlining the construction of returns and sorting variables, before elaborating on the construction of portfolio sorts and dependent variables and finally explaining the construction of factor premia and independent variables.

4.5 Explanation of Sorting Variables

The following subsection describes the sorting variables and their construction as well as the construction of additional variables necessary for the estimation of the models of interest. First, the general time series variables are explained and subsequently the variables to construct the factors used in the three-, four- and five-factor models are explained.

Individual Stock Returns

To calculate the underlying returns of each stock, the Total Return Index (*RI*) available on *Datastream* is used as it accounts for stock splits, dividends and other changes to the stocks. The calculation of the returns is straightforward as it is the change of the *RI* over the last month:

$$r_{t,i} = \frac{RI_{i,t}}{RI_{i,t-1}} - 1$$

Portfolio Returns

To calculate the monthly return of a portfolio $r_{t,pi}$, the monthly return r_i of each stock *n* within a portfolio is weighted according to its relative portfolio *Floating MV*

$$r_{t,pi} = w_{t,ni} * r_{t,Stock_{ni}},$$

where the weights are given by

$$w_{t,ni} = \frac{MV_{t-1[Stock]}}{MV_{t-1[Portfolio]}},$$

where $MV_{Stockn,i}$ is the *Floating MV* of a Stock and $MV_{Portfolio}$[49] is the total *Floating MV* of the portfolio. The individual stocks *n* in each month *t* are weighted by the value weights calculated at the end of month *t-1*.

[49] $MV_{Portfolio_i} = [MV_{Stock1,i} + \cdots + MV_{Stockn,i}]$

Floating MV is used instead of *Total MV* to construct the value weights of the portfolios for two reasons. First, the observed returns in the market stem from *Floating Shares*. Stocks that are non-floating do not deliver observable returns in the Chinese market. *Datastream* starts reporting Total Return Index (*RI*) as soon as *Floating Shares* are available in the market. By weighting with *Total Shares* instead of *Floating Shares*, returns which stem from companies having a small amount of *Floating shares* but a large number of *Total Shares* would overstate the importance of this returns to the composition of the Portfolio returns. Second, non-floating share prices are not publicly available and mainly based on the underlying book value of equity. As they are only allowed to trade between large state-linked institutional investors, they are subject to different pricing mechanisms and are therefore traded non-publicly in over-the-counter (*OTC*) deals between these institutions.

Right Timing of the weights is crucial for the precision of the returns. A different timing, using the *Floating MV* in each month *t* to calculate the weights for the same month would lead to an overstatement of portfolio returns, as the weights would already account for the increased *MV* during the month[50]. Therefore, it is appropriate to calculate the weights using the information available at the start of each period, hence *Floating MV* of the end of the previous period is used.

$$\overline{r_{pi}} = \frac{1}{T}\sum_{t=1}^{T} r_{t,pi}$$

Excess Returns

The excess return is the return of a portfolio after the risk-free rate in the market is subtracted.

$$excess\ return_{pi} = \overline{r_{pi}} - r_f$$

They are used to evaluate the performance of every single portfolio in the dependent variable portfolio sorts. Furthermore, excess returns are used for the construction of the market factor.

Market Returns

Another variable that needs to be constructed is the market return (R_M). In developed markets, it is common to use a major index with a sufficient amount of constituents as a proxy for the returns of the market[51]. In the absence of available Chinese index data, the market return is constructed from the full set of stocks available in the stock markets included in our data sample (Shanghai Stock Exchange

[50] $w_t = \frac{MV_{t[Stock]}}{MV_{t[Portfolio]}}$

[51] For the US market S&P500 or Russel 2000 are typically used as market proxies

and Shenzhen Stock Exchange). This is done by calculating the monthly value weighted average return of each stock. First monthly market weights for each stock are calculated. Therefore, the *Floating MV* of each stock is divided by the total Sum of the *Floating MV* per month. The achieved weights are multiplied with the monthly returns of each stock. These return fractions are summed up over the entire market.

$$w_{t,i} = \frac{Float\ MV_{t,i}}{Float\ MV_{t,Total}}$$

$$r_{t,M} = \sum_{i=1}^{t} w_{t,i} * r_{t,i}$$

Overall, there exists a wide range of methods and factors to construct *factor*- or *APT-Models.* As already mentioned above this thesis mainly focuses on the Fama and French approach (1992, 1993). As the Chinese market only opened up recently and specific accounting information on listed companies is still not entirely available this thesis focuses rather based on the simpler and well documented 3- and 4-factor models than on more complex approaches.

Size

First, the Company Size variable (*Size*) is constructed as the Market Value of Equity (*MV*) at the end of June each year *t*[52]. *Size* is kept constant from June year t until next year's June *t+1* where it is re-estimated using the corresponding *MV* in June *t+1.*

$$Size_t = MV_{t,June}$$

When constructing the *Size* measure, it is necessary to decide whether to use total value or floating market value. This is not an issue in most other markets since total size usually is well measured by the tradable market capitalization, but due to the history of large amounts non-listed common equity in China, it is relevant for this thesis. Xu and Zhang (2014) investigate the influence of different construction methods of the independent variables in the Fama and French three-factor model, and one of their main findings is that size should include both floating and non-floating shares. Despite their recommendation, most of the recent factor literature written on the Chinese market only uses floating market value for the size measure.

52 We follow the Fama-French argumentation and construct the Size variable at the end of June each year because it is assumed that until than all public announcements and singles stemming from annual reports are reflected in the equity prices

This thesis follows the recommendation of Xu and Zhang (2014), arguing that the definition of size is best done on the basis of all equity in a firm, not only a subset of the equity. In the case that only floating market value is used, the size of a firm could be misleading, and this would disrupt the measurement of the size effect. This implicitly assumes that the market value of non-floating shares would be the same as the market value of floating shares if they were publicly traded, which seems like a reasonable assumption given the equal claim to equity of the different share classes[53]. However, the influence of changing total market value with floating market value will be shown and discussed in the robustness tests in section 7.2. It should be noted that total *Size* as constructed here only is used to define *Size* to divide firms into different portfolios for both the dependent and independent variable construction. Value weighting of returns is done by the floating market value, as explained in deeper detail in the specific part on returns.

Book-to-market

Second, following Fama and French (1993) the Book-to-Market variable (*B/M*) is constructed as the Book Value of Equity (*BE*) divided by the Market Value of Equity. For the same reasons applying to the *Size* variable Market Value of Equity is defined as *Total MV*. At the end of December in year *t-1,* we measure *Size* and *BE. B/M* is calculated (as the *Size* variable) every June and kept constant on its June level in year *t* until next year's June *t+1* where it is re-estimated using the corresponding *B/M* in June *t+1* (calculated from the *Size* and *BE* values in end of December in year t).

$$B/M_t = \frac{BE_{t-1,December}}{MV_{t-1,December}}$$

However, similar to the size measure, the *B/M* measure is affected by the particular Chinese share structure. Xu and Zhang (2014) recommend adjusting the traditional book value of equity by a measure constructed as the book value of equity divided by the total number of shares outstanding to obtain the book value per A-share. Then the book value per A-share is divided by the price per A-share. That is the book value of equity divided by total market value on an aggregated level. This adjustment is made because more shareholders than just the A-shareholders have a claim to the total book value of equity. Therefore, it is necessary to adjust the total book value to the book value A-shareholders have a claim to. This adjustment is made in most of the recent Chinese three-factor model research[54].

53 Xu and Zhang (2014)
54 See for example Xu and Zhang (2014), Chen et al. (2015) and Hu et al. (2018)

Momentum

Third, the *Momentum* variable is constructed as an average over a predefined period of past returns. The *Momentum* “variable” can be constructed following different formation mechanisms.

1.) In earlier literature, it was common to construct the *Momentum* factor based on a *t-j* observation window, holding a portfolio based on this factor from t to *t+k* and selling the entire portfolio at the end of *t+k*. This approach was introduced by Titman and Jegadeesh (1993) and tested for several periods of creation and holding. In their sample on the US-market, they find the optimal period of formation and holding to be j=6 (formation) and k=6 (holding). Following Titman and Jegadeeshs approach, the *Momentum* variable in month *t* is formed as the simple average over the months *t-1* to *t-6.* The variable is kept constant for the next six months (*t* until *t+5*). This implies half yearly rebalancing of the variable instead of yearly as done before.

$$momentum_{tit} = \frac{1}{6} * \sum_{i=-1}^{T} return_i \qquad with\, T = -6$$

The variable is static for six months, and the window is moved every half year to estimate a new static *Momentum* variable. This is the main difference compared to the second *Momentum* variable.

2.) *Later on, a “new” Momentum variable* was introduced by Carhart (1997) in his paper on mutual fund performance. Carhart defines the *Momentum* factor as the simple average over a stock’s past year returns. Following his approach, a *Momentum* factor is built every month on a rolling base. In each month *t-1*, the simple average over the last 12 months skipping the month of formation (*t-1*) is calculated. We define the *Momentum* in each month as the past mean return of each stock:

$$past\ mean_{t,i} = \frac{1}{11} * (r_{t-2,i} + \cdots + r_{t-12,i})$$

This effectively results in a Cumulative Moving Average (*MA*) that drops the earliest month each time it is shifted and takes in the second latest month. This implies a construction period of 12-1 months and a holding/ rebalancing period of 1 month and leads to 12 different Momemtum_CAR values per year per stock. Only the second method is investigated in the analysis, but the robustness of the momentum findings are tested in Section 7.2 by also following the first estimation method.

Investment

The next two variables are input factors in the relatively new five-factor model that was developed over the course of the last five years. *Inv* is constructed as the change in the Total Assets (*TA*) of a company. In each June of year *t,* the *TA* value at the end of December of year *t-2* is subtracted from the end of December *TA* value in t-1. This absolute change in Total Asset, the total investment over the last year *t-1* is divided by the *TA* value of December t-2. This fraction displays the percentage change of *TA* during the last year *t-1.* Like *Size* and *B/M,* this variable is re-estimated in June of each year.

$$Inv_t = \frac{TA_{t-1} - TA_{t-2}}{TA_{t-2}}$$

Operational Profit

The *OP* variable is constructed as a company's annual revenues (*Revenues*) measured at the end of December in year *t-1* minus Cost of Goods Sold (*CoGs*), Interest Expenses (*Interest*), Depreciation and Selling, General, and Administrative Expenses (*SAG*) all divided by *BE* at end of December in year *t-1*.

$$OP_t = \frac{Revenues_{t-1} - CoGs_{t-1} - \ Depreciation_{t-1} - SGA_{t-1}}{BE_{t-1}}$$

Depreciation is subtracted from the *Revenues*, as it is not subtracted from the *Revenues* obtained from *Datastream*. Moreover, the *Interest* is not subtracted due to their seldom availability[55]. As before the *OP* variable is recalculated at the end of June in each year *t+1*.

Dividend-to-Price & Earnings-to-Price

Dividend-Price (*D/P*) and Earnings-Price (*E/P*) are two additional variables. They are mainly used for descriptive statistics. *D/P* is the simple monthly average of the *D/P* per month calculated over the year.

$$\overline{D/P}_{montly} = \frac{1}{12}\sum_{t=1}^{12} D/P_{t,i}.$$

E/P is also calculated as the simple monthly average of the *E/P* calculated over the year.

$$\overline{E/P}_{montly} = \frac{1}{12}\sum_{t=1}^{12} E/P_{t,i}.$$

[55] Interest expenses where available for 150 of ca. 3200 companies that amounts to less than five percent. We therefore decide to drop this variable. Even in the Fama and French paper their use seems to be a bit arbitrary as specified by Fama and French that they subtract them when available

The Chinese risk-free rate r_f is used to calculate monthly excess returns.

After describing the underlying variables of the models, the thesis proceeds with the construction of portfolio sorts.

4.6 Portfolio sort/ generation

In the following section portfolio sorting techniques are introduced and explained. Sorting specific companies/assets into portfolios is the base for the creation of all dependent and independent variables, and it is one of the crucial steps of the analysis.

General Portfolio construction

As the focus of this thesis in line with the general focus in academia lies on the influence of systematic risk on the overall development of returns and not on the performance of specific firms/ assets, idiosyncratic risk should be minimized. To diversify the non-systematic risk appropriately, sufficiently large portfolios are formed. Portfolio formation provides two advantages – first, the idiosyncratic risk is minimized and second a factor based sort of the dependent variable (in asset pricing typically returns or excess returns) provides first insights in the development of the dependent variable over factor characteristics.

The disadvantage of portfolio formation is that the resulting model cannot explain the returns of single stocks ($E[R_p] - r_f$ and not $E[R_i] - r_f$ is observed). A second problem that might occur, especially in emerging markets, is the fact that a large matrix structure (*nXm*) might lead to effectively low diversification due to the absence of sufficient amounts of stock in each portfolio. For a 5X5 Portfolio matrix (effectively 25 double-sorted portfolios) that should fulfill sufficient diversification standards at least 25*30 = 750 stocks are needed.

In their initial factor model papers, Fama and French (1992 and 1993) form dependent variable portfolios on the market value of firms (*Size*) and on the book value of equity divided by the market value of firms - book-to-market (*B/M*).

Univariate Sort

To investigate the general behavior of factor sorted returns and to obtain a first view on trends within excess returns, univariately sorted portfolios are constructed as follows. At the end of June each year, every stock in the market is individually ranked on *Size*, *B/M*, *Momentum*[56], *Inv,* and *OP*. From these rankings, ten univariate portfolios each are constructed. The breakpoints for these ten portfolios are

[56] Momentum build after the Momentum variable Carhart introduced

set by each variables market deciles. Value-weighted monthly returns for each of the ten variable sorted portfolios *i* are constructed. The averages over these value-weighted monthly returns build the first simple factor premia, whereby the simple “factors” are only based on one-dimensional sorts in comparison to the later on constructed “real” factor premium estimates. The simple factor returns provide a first indication of which sorts might be more profitable than others and therefore which factors might be more successful than others. The univariate sorts should neither be characterized as dependent variable sorts nor as independent variable sorts. They are only constructed to provide a first overview on variable based portfolio sorts.

Dependent Variable

After the univariate sorting provide the first overview on the general behavior of different variable based sorts, the dependent variable portfolios for the later performed regressions are constructed following Fama and French´s (1993) multivariate sorting approach.

Multivariate Sort

In a next step, multivariate dependent variable portfolios are constructed. Fama and French (1992 and 1993) do this by double sorting excess returns on the factors of interest. As mentioned earlier portfolio formation provides the advantage that idiosyncratic risk is highly diversified. An additional advantage stemming from multivariate sorting is that it helps to distinguish the specific return behavior of one variable from the influence of the other variable (e.g. *Size* vs. *B/M*). The wide range of excess returns produced by this sorting technique offers a suitable base for testing asset-pricing models (Fama and French, 1993). Subsequently, four double-sorted portfolios and three triple sorted portfolios are introduced.

Size-*B/M* Portfolios

First, 25 portfolios (5x5) on *Size* and *B/M* are constructed. Fama and French investigate the relation between these two risk factors and average excess returns in their initial paper on *Size* and *B/M* (1992). *Size* and *B/M* were introduced into the Asset pricing world as a proxy for underlying economic fundamentals. As mentioned before firms with high *B/M* ratios tend to have high returns, whereas firms with low *B/M* tend to have low returns. Furthermore, smaller firms (small *Size*) tend to have higher returns, and bigger firms (big *Size*) tend to have lower returns. Inspired by these findings the above-mentioned double sorting is motivated. For Fama and French (1993) this set-up leads to a pronounced size and value pattern depicted in the 25 portfolio excess returns. This is illustrated by

$$\begin{pmatrix} p_{S1} \\ p_{S2} \\ p_{S3} \\ p_{S4} \\ p_{S5} \end{pmatrix} \times \begin{pmatrix} p_{B1} & p_{B2} & p_{B3} & p_{B4} & p_{B5} \end{pmatrix} = \begin{pmatrix} p_{SB11} & \cdots & \cdots & \cdots & p_{SB15} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ p_{SB51} & \cdots & \cdots & \cdots & p_{SB55} \end{pmatrix},$$

where p_{Si} are the five *Size* sorted portfolios, p_{Bj} are the five *B/M* sorted portfolios and p_{SBij} are the *Size*–*B/M* sorted portfolios

More particular, in end of June each year *t Size* and *B/M* are independently ranked and sorted into five *Size* quantiles and five *B/M* quantiles. Following previous literature on the Chinese market (e.g. Lin (2017)), the borders of the *Size* and *B/M* groups are defined as the simple quintile breakpoints of the entire market[57]. More generally, breakpoints in this study are consistently formed on quantiles of the entire market. The intersects of the independently created *Size* and *B/M* groups are used to form the 25 *Size-B/M* portfolios. Stock with negative *B/M* are excluded.

Value weighted monthly excess returns for each of these 25 portfolios are created from July of year *t* to June of year *t+1*. The weights and returns are calculated as in the univariate case. In every month *t,* the excess return $r\text{-}r_f$ of each stock *n* within each of the 25 portfolios is weighted according to its relative portfolio Market Value (*MV*). The value-weighted monthly excess returns are used as dependent variable in the further analysis.

To display isolated overall average patterns for each of these 25 excess returns, average monthly excess returns over the entire period are calculated.

Size-Momentum Portfolios

Moving further in time additional portfolio sorting techniques were used. After the introduction of *Momentum* into factor models (Carhart (1997)) alternative sorting on *Size* and *Momentum* evolved.

The formation of the excess return portfolios closely follows the formation of the *Size-B/M* Portfolios. At the end of June of each year *t*, *Size* and *Momentum* are independently ranked and sorted into five *Size* quantiles and five *Momentum* quantiles. The intersects of the independently created *Size* and

[57] Unlike studies on the US market, where the break points are solely defined as the quintiles on the NYSE

Momentum groups are used to form the 25 *Size-Momentum* portfolios. Value-weighted monthly excess returns are calculated as described for *size-B/M* above.

Size-Inv & Size-OP Portfolios

In one of their latest papers (2015), they introduce a five-factor based model. Here they partly replace the *B/M* factor by two newly introduced variables, *Inv* and *OP*. Again, dependent variable portfolios of value-weighted excess returns are formed. *Size-Inv* Portfolios are formed in the same manner as *Size-B/M.* At the end of June of each year *t*, *Size* and *Inv* are independently ranked and sorted into five *Size* quantiles and five *Inv* quantiles. The intersects of the independently created *Size* and *Inv* groups are used to form the 25 *Size-Inv* portfolios. Value-weighted monthly excess returns are calculated as described for *Size-B/M* above. For the *Size-OP* portfolios, the same procedure as for the *Size-Inv* portfolios applies.

Summary Statistics of Dependent Variable sorts

After creating the different dependent variable sorts, summary statistics among others containing the "Average of annual number of firms in portfolio," the "Average of annual percentage of market value in portfolio" and the "Average of annual *B/M* ratios for portfolio" are created.

The average excess returns of each of the portfolios in every multivariate sorted matrix are calculated to assess the performance of each of these portfolios. Excess return matrices are used by Fama and French (1993) to evaluate which variable combinations lead to systematic high returns and are therefore the basis of the factors/ independent variables of the model. Furthermore, t-statistic matrices and standard deviation matrices are calculated to assess the excess return matrices.

Factor Creation / Independent Variable sorts

A core input to factor model regressions, as with every empirical model are the independent variables. The independent variables, in this case the factors or factor mimicking portfolios, are used to explain the behavior of the dependent variable, in this case the excess return portfolios. Factor mimicking portfolios are constructed as self-financing zero-investments and are therefore always long short combinations[58]. As research has moved on, standard models have been expanded, and additional factors have been added.

In the following section, the construction of the independent variables is explained.

The Market factor

The market factor is used as an independent variable in most asset pricing models. In the CAPM, it is the only risk factor used for estimating the market beta. The Market Return (R_M) constructed as shown in the variable part of the methodology. The market factor is obtained by subtracting the risk-free rate from the R_M. The market factor is different from the rest of the factors as it is not a self-financing long-short combination. It is just as little a combination of two variables. Nevertheless, the market factor is financed by shorting the risk-free rate.

The Size factor

Following Fama and French (1993), 2x3 independent variable portfolios on *Size-B/M* are formed. Two portfolios on *Size* and three portfolios on *B/M* are created independently at the end of June each year *t*. The breakpoint for the *Size* is at the median of the sample. The breakpoints for the *B/M* sorting are defined such that the lowest ranked 30% stocks will be part of the low-group, the medium ranked 40% stocks will be part of the medium-group, and the highest ranked 30% stocks will be part of the high-group. From the intercepts of these two sorts the six portfolios S/L, S/M, SH, B/L, *B/M*, B/H are created. To construct the risk-mimicking portfolio, averages over stocks with the same portfolio characteristics are built.

For the Size mimicking factor portfolio, the Small (*S*) factor is calculated as:

$$S_t = \frac{{r_t}^{S/L} + {r_t}^{S/M} + {r_t}^{S/H}}{3}$$

and the Big (*B*) factor is calculated as:

$$B_t = \frac{{r_t}^{B/L} + {r_t}^{B/M} + {r_t}^{B/H}}{3}$$

[58] In general, there exist also factor mimicking portfolios which are not strictly self-financing. In the scope of this thesis all factor mimicking portfolios are self-financing.

where $r_t^{X/Y}$ is the monthly value weighted[59] return on a portfolio formed on X/Y

Finally, the Small minus Big (*SMB*) value weighted mimicking factor Portfolio is created as:

$$SMB_t = S_t - B_t$$

Fama and French (1993) argue that this factor should be free of any *B/M* influences due to its construction. Furthermore, they argue that due to the value-weight set-up the variance is minimized as well as capturing return behaviors that are linked to a more realistic investment opportunity.

The Value *factor*

The breakpoints for the *Value* factor for *Size* and *B/M* are the same as for the *Size* factor.

For the *B/M* mimicking factor portfolio, the High (*H*) factor is calculated as:

$$H = \frac{r_t^{S/H} + r_t^{B/H}}{2}$$

and the Low (*L*) factor is calculated as:

$$L = \frac{r_t^{S/L} + r_t^{B/L}}{2}$$

where $r_t^{X/Y}$ is the monthly value weighted return on a portfolio formed on X/Y

Finally, the High minus Low (HML) value weighted mimicking factor Portfolio is created as:

$$HML_t = H_t - L_t$$

Similar to the *Size* factor, Fama and French (1993) argue that this factor should free of any *Size* influences due to its construction.

The Momentum *factor*

As mentioned earlier there are two ways to the *Momentum* factor in the literature. Fixed formation and execution periods of Momentum Portfolios and the "rolling," Moving Average approach used by Carhart and Fama and French.

[59] As with the univariate sorted portfolios and the dependent variable portfolios, value weights are calculated with Floating Market value

First, for the construction of the Titman *Momentum* factor monthly returns on the winner portfolio are calculated from the top-90% decile, and the returns on the looser portfolio are calculated on the bottom-10% decile. The *W*, *L* and *WML* portfolios are calculated as follows:

$$W_t = \sum_{i=1}^{n} r_{i,t}{}^{high90\%}$$

$$L_t = \sum_{i=1}^{n} r_{i,t}{}^{low10\%}$$

$$WML_t = W_t - L_t$$

Second, following Carhart's (1997) and Fama and French's approach (2004) the *Momentum* factor is calculated on a yearly moving average base[60]. As illustrated before the past mean variable is calculated on a rolling base and portfolios, although constructed on a yearly base are only held for one period. Like *B/M* we rank our yearly average past returns into three different portfolios. The breakpoints for the *Momentum* sorting are defined such that the lowest ranked 30% stocks will be part of the Looser-group, the medium ranked 40% stocks will be part of the Neutral-group, and the highest ranked 30% stocks will be part of the Winner-group. From the intercepts of the *Momentum* portfolios and the two *Size* portfolios sorts the six portfolios S/L, S/N, SW, B/L, B/N, B/W are created.

To construct the risk-mimicking portfolio, averages over stocks with the same portfolio characteristics are built. For the *Momentum* mimicking factor portfolio, the Winner (*Wi*) factor is calculated as:

$$Wi = \frac{r_t{}^{S/W} + r_t{}^{B/W}}{2}$$

and the Looser (*Lo*) factor is calculated as:

$$Lo = \frac{r_t{}^{S/L} + r_t{}^{B/L}}{2}$$

where $r_t{}^{X/Y}$ is the monthly value weighted return on a portfolio formed on X/Y

Finally, the High minus Low (*WML*) value weighted mimicking factor portfolio is created as:

$$WML_t = W_t - L_t$$

[60] As before a one year creation period essentially means the average of the past years returns skipping the month of factor creation (t-1)

Five-factor Model Independent Variables

Following Fama and French (2015) 2x3 independent variable portfolios on *Size-OP and Size-Inv* are formed. Similar to the dependent variable portfolio creation, at the end of June of each year *t*, two portfolios on *Size* and three portfolios on *OP* and *Inv* are created independently. The breakpoint for the *Size* is at the median of the sample. The breakpoints for the *OP* and *Inv* sorting are defined such that the lowest ranked 30% stocks will be part of the Low-group, the medium ranked 40% stocks will be part of the Medium-group, and the highest ranked 30% stocks will be part of the High-group.

The OP *factor*

From the intercepts of the sorts on *Size-Op* six portfolios S/W, S/M, S/R, B/W, *B/M*, B/R are created. To construct the risk-mimicking portfolio, averages over stocks with the same portfolio characteristics are built. Value weighted portfolio returns are calculated for S/R, B/R, S/W and B/W and the Robust minus Weak (*RMW*) value weighted mimicking risk factor portfolio is calculated as:

$$RMW_t = \frac{{r_t}^{S/R} + {r_t}^{B/R}}{2} - \frac{{r_t}^{S/W} + {r_t}^{B/W}}{2}$$

The Inv *factor*

From the intercepts of the sorts on *Size-Inv* six portfolios S/C, S/M, S/A, B/C, *B/M*, B/A are created. To construct the risk-mimicking portfolio, averages over stocks with the same portfolio characteristics are built. Value weighted portfolio returns are calculated for S/C, B/C, S/A, and B/A and the Conservative minus aggressive (*CMA*) value weighted mimicking risk factor portfolio is calculated as:

$$CMA_t = \frac{{r_t}^{S/C} + {r_t}^{B/C}}{2} - \frac{{r_t}^{S/A} + {r_t}^{S/A}}{2}$$

Average Factor Premia

To assess the overall direction, size and significance of the factor, its average premium, standard deviation, and t-statistic are calculated. The t-statistic verifies whether the premium is significantly different from zero or not. To be considered as an actual premium the t-statistic has to deliver a significant value. Furthermore, correlations between the factors are calculated to assess potential multicollinearity.

4.7 The estimated Models

After calculating all necessary input variables and performing all relevant sorting, the actual models are estimated. As briefly mentioned before, the models are time series estimates where the dependent variable is an excess return matrix, and the independent variables are the risk factors. This leads to the following generalized model:

$$\begin{bmatrix} p_{11} & \cdots & \cdots & \cdots & p_{1m} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ p_{n1} & \cdots & \cdots & \cdots & p_{nm} \end{bmatrix} = \alpha_i + \beta_{i,1} Factor_1 + \beta_{i,2}\ Factor_2 + \varepsilon_i$$

, where *nxm* is the size of the dependent variable matrix, α_i is the intercept, $\beta_i$is the factor exposure and ε_i is the error term

Following Fama and French (1993) each portfolio of the excess return matrix is regressed on the independent variables in a separate regression. The resulting intercepts and factor exposures as we, as well as their t-values, are collected in matrices. Furthermore, the adjusted R-squares of all regressions are collected in a matrix.

The intercepts and R-squares are collected to assess the overall fit of the estimated models. Low overall intercept values imply that the factors explain most of the returns present. High adjusted R-square values imply that most of the variation in the dependent variable is explained by the model[61].

T-values are used to assess the statistical significance of each factor and intersect.

More specific, the following models are estimated:

<u>CAPM</u>

$$r_{pi,t} - r_{f,t} = \alpha_i + \beta_i\left[r_{M,t} - r_{f,t}\right] + \varepsilon_i, \qquad i \ = \ 1, \dots, n \: and \: t = 1, \dots, T$$

<u>Three-factor model</u>

$$r_{pi,t} - r_{f,t} = \alpha_i + \ \beta_{i,m}\left(E\left[r_{M,t}\right] - r_{t,f}\right) + \ \beta_{i;SMB} SMB + \ \beta_{i,HML} HML + e_i. \qquad i \ = \ 1, \dots, n \: and \: t = 1, \dots, T$$

---

[61] Especially in a time series set-up, the R-squared measure tends to suggest a high explanatory power due to the fact that aggregate dependent variables are often reported in aggregated form (in the our case we try to explain monthly access returns of portfolios this is easier than explaining daily returns of single companies).

Four-factor model

$$r_{pi,t} - r_{f,t} = \alpha_i + \beta_{i,M}\left(E[r_M] - r_{t,f}\right) + \beta_{i;SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,WML}WML_t + e_i, \qquad i = 1, \dots, n \; and \; t = 1, \dots, T$$

Five-factor model

$$r_{pi,t} - r_{f,t} = \alpha_i + \beta_{i,M}\left(E[r_M] - r_{t,f}\right) + \beta_{i;SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,RMW}RMW_t + \beta_{i,CMA}CMA_t + e_i, \qquad i = 1, \dots, n \; and \; t = 1, \dots, T$$

Depending on the specified model, the monthly excess returns of the criteria sorted portfolios *Size-B/M*, *Size-Momentum*, *Size-OP*, *Size-Inv* are regressed on the market-, size-, value-, momentum-, profitability- and investment-factors.

To avoid misinterpretations or overconfidence issues, additional goodness of fit measures are introduced. The next section elaborates on these mainly alpha based measures and their interpretation.

4.8 Measures of fit:

All the above mentioned measures of fit are only able to test for the goodness of fit of each of the single regressions of the dependent variable matrix but not for the goodness of fit of the overall model. As it is challenging to judge and compare the models to each other when looking at every single part of the output matrices an overall measure of fit is needed which provides a single test statistic, R squared, and p-value one can compare the models directly to each other.

GRS Estimations:

One of the most commonly used measures of fit in modern asset pricing tests is the Gibbons Ross Shanken (*GRS*) test. The *GRS* test works similar to an *F-test*. It tests the joint hypothesis that all alphas are zero. Achieving joint zero alphas is the main condition for a mean-variance efficient portfolio.

$$H_0: \alpha_i = 0 \;\; i = 1, \dots, N$$

The test statistic is defined by:

$$W_T = \frac{T - N - K}{N} * \left[1 + \overline{f'}\widehat{\Omega}^{-1}\overline{f}\right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \;\sim F_{N,T-N-K}$$

Where T is the number of months included in the time series, N is the number of portfolios, T-N-K specifies the number of degrees of Freedom, $\overline{f}$ is a mean factor vector, $\widehat{\Omega}$ the variance of the factor and $\hat{\Sigma}$ is the residual covariance matrix

The displayed test statistic assumes normally distributed standard errors.

If the calculated test statistic is large the H_0 will be rejected with a high probability. In comparison to a "regular" F-statistic, the goal of the *GRS* test is to not reject the H_0. Therefore a low test statistic is necessary. Similar to the F-test the corresponding significance levels depend on the degrees of freedom of each test. In general *GRS* tests which fail to reject the H_0 on at least a 5% level are desirable.

Fama and French (2012, 2014, 2016) use *GRS* values to compare their estimated models. The lower the test-statistic becomes, the better the model fits. Fama and French (2012) report a clear pattern of falling *GRS* scores when moving from global *CAPM* tests to tests of the 3-factor and 4-factor model.

By introducing a method to measure the overall fit of the single models as well as allowing for cross-model comparison, the methodology has provided all the necessary tools to perform the factor analysis. But first, after elaborating on the methodical approach taken to estimate the models of the thesis, the next section the necessary model input data.

Section 5 – Data

This section deals with gathering and treatment of data used as input for the analysis in the thesis. The data used in this thesis is gathered from *"Thomas Reuters Datastream"* for market data and *"Thomas Reuters Worldscope"* for accounting information. Equity market data from *Datastream* and fundamental data from *Worldscope* is sourced directly from exchanges, international suppliers and published reports[62].

First, the subsection on data gathering describes the data needed for the main analysis and how it is obtained before the following subsection describes a data cleaning procedure that builds on the works of Ince and Porter (2006) and Schmidt et al. (2017). This section relies heavily on the data cleaning techniques presented in the two papers above, since assessing the quality of *Datastream* data is not the main focus of this thesis.

[62] http://share.thomsonreuters.com/assets/newsletters/ssr/Datastream.pdf

5.1 Data gathering

This part shortly describes the data necessary for the later analysis, as well as justifies the choice of input if it differs from the common approach. As elaborated on in the methodology, the inputs necessary for the *CAPM*, the three-factor model, and the four-factor model are returns, a risk-free rate, market capitalization and the book-to-market ratio. In addition to this, a profitability factor and an investment factor are necessary for the five-factor model. How each of the components is obtained is therefore described in separate sections below, after the overall stocks and period of investigation are described.

Constituent lists

As pointed out in section 2, this thesis concerns itself with the Chinese, floating A-share market. The constituent list therefore optimally contains all A-share listings on Shanghai and Shenzhen Stock Exchange. The data is obtained by downloading data for firms in the *Datastream* lists "Shanghai SE A-Share" (1408 constituents) and "Shenzhen SE" (2094 constituents), in total 3502 constituents as of 28th of April. Another possible approach previously taken by Cheung et al. (2015) could be to use the MSCI China A-Share index, but this is unfortunately not feasible for this thesis due to the lack of available MSCI data. Using the MSCI index provides a better description of the currently investable universe in China, but the trend of a more open China could make the description of a wider equity universe equally useful going forward. Furthermore, Fama and French (1998) state that preliminary tests they perform indicate that a database only including large stocks does not allow for meaningful tests of the size effect.

Following Fama and French (1993) and others, financial institutions[63] are excluded from the sample due to their different leverage mechanisms. The summary statistics, as well as regressions on a sample including the financial firms, are included in the Appendix.

Period of investigation:

The purpose of this thesis is to describe the characteristics of the Chinese A-share market today, and the characteristics in the period investigated should be of a similar nature. Shanghai and Shenzhen Stock Exchange were both founded and started operating in 1990. Hence that represents the longest

[63] Such as Banks, Insurances, Asset and Capital Management firms and other Financial firms

possible period of investigation. It is safe to say that the Chinese market has experienced rapid development and frequent regulatory changes over the last decades. That makes it challenging to pick a starting point for the empirical analysis, balancing the need for a market that throughout the period is similar to today's market and the need for a longer period to perform any statistically meaningful analysis. It has been argued that not even sample periods of 20 years are long enough for ex-post returns to proxy for expectations. However, in terms of similarity to today's market a point in time that stands out is the years between 2004 and 2006, with easing of listing standards through the SME Board in 2004, a considerable amount of non-floating shares being floated between 2005 and 2007, and regulatory authorities allowing short selling and margin trading in 2006.

The authors are aware of the limits the short period imposes for solid conclusions to be drawn, but the changes in the market structure are seen as so big that it is necessary to use a shorter period. However, the results must of course be seen in this context and interpreted with caution. The period will also be extended back to 1996 in the robustness test to look at a larger picture. The main part of the analysis will investigate stock returns in the period between July 2006 and June 2017. Most of the important regulations have been present since 2006, and the period of 132 months still allows for meaningful statistical analysis.

5.2 Inputs to the models

Returns:

For the construction of portfolios, return data on individual Chinese equities is needed. *Datastream's Total Return Index* (datatype *RI*) returns the overall development of capital gains and dividend payments. That is the value of an investment where all dividends are reinvested in the stock at the moment they are paid out. The monthly return for stock *i* at time *t* is then calculated as

$$r_{t,i} = \frac{RI_{i,t}}{RI_{i,t-1}} - 1$$

The monthly returns of the constructed portfolios are value-weighted by their monthly floating market value, constructed as the product of the number of A-shares (datatype *NOSH*) and the unadjusted price (*UP*).

Risk-free rate:

The risk-free rate should represent the return of a risk-free investment, and the return on what is assumed to be a risk-free asset is used as a proxy for this. In addition, constructing *APT* consistent portfolios includes financing the market portfolio by shorting the risk-free rate. Thus, the risk-free rate should be tradable. The conventional approach in the literature is to use a short-term Treasury bill rate, as Fama & French (1993) also do. However, treasury bills do not have such a long history in China. Most of the bonds with different maturity were issued at some point between 2000 and 2010[64]. *Datastream* for example provides data for the 1-year and 3-month government bonds from May 2003.

The approach that is taken in much of the factor literature[65] on the Chinese market is therefore to use a time deposit rate. As Allen et al. (2017) point out, the time deposit rate can be used because the government owns the majority of all large banks. Hence, deposit rates are effectively risk-free rates. The first choice that has to be made is which one of a time deposit rate and a government bond rate that should be used. Since the period of investigation in this paper starts in 2006, as further elaborated on in the previous section, a government bond could be used. However, longer periods will be used in the robustness section at the end of the paper, and the Chinese time deposit rate is used for reasons of comparison. The time deposit rate is nevertheless switched with the 3-month central bank bill in the robustness section to check if the choice of risk-free rate makes a significant difference.

A second choice that should be made is the horizon of the deposit rate. Both the 1-year rate and the 3-month rate are commonly chosen in the literature, with Drew et al. (2003) and Gan et al. (2013) as examples of the former, and Zhang and Xu (2014) as an example of the latter. Zhang and Xu motivate their choice of the risk-free rate by arguing that the 3-month rate better matches the monthly returns under observation, and this thesis follows their approach and goes with the 3-month time deposit rate (datatype *CHSRW3M*).

[64] See Bai et al. (2013) for further details on the Chinese government bond market.
[65] See for example Xu and Zhang (2014) and Cheung et al. (2015)

Figure 4 - Development of the time deposit rates

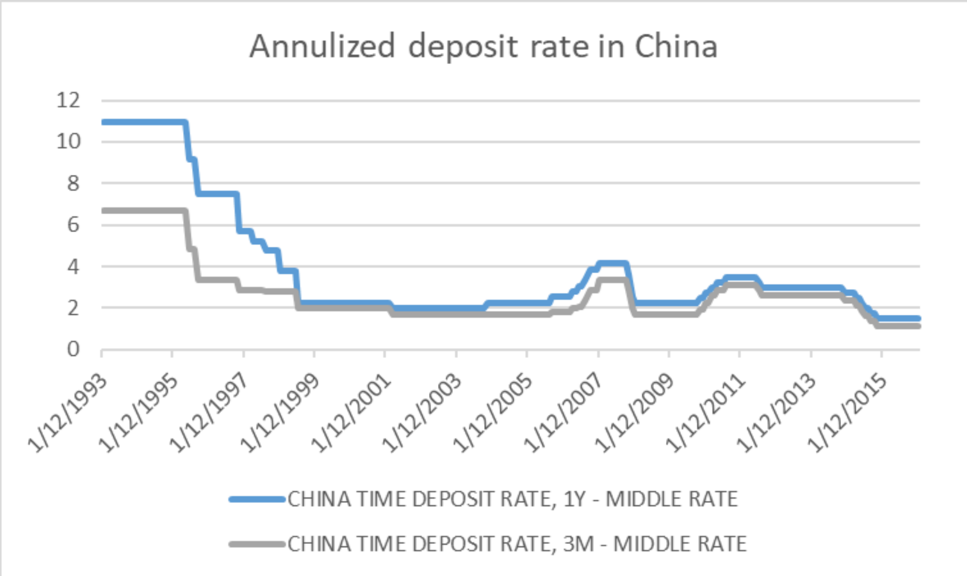


Figure 4 shows the development of the 1-year and 3-month time deposit rates from 1993 to 2016. It shows that the two rates have had a similar development. However, the 3-month rate has consistently been lower than the 1-year rate, and it has had less significant spikes. *Datastream* states interest rates in percentages on an annual basis, and this is also the rates shown in Figure 4. To obtain monthly returns for the analysis, the 3-month time deposit rate at time t is divided by 100 and scaled as

$$(1 + rf_t)^{1/12} - 1$$

Market capitalization

Schmidt et al. (2017) measure market value either as the market value provided by *Datastream* (datatype *MV*) or as the product of unadjusted price (datatype *UP*) and number of shares (datatype *NOSH*). *MV* is calculated as the product of *UP* and *NOSH* by *Datastream*, so the latter calculation is only used to supplement the former raw data from *Datastream* if it for some reason is missing. *Datastream* defines *NOSH* as ”*the total number of ordinary shares that represent the capital of the company*.” Furthermore, they state that this datatype holds each equity issue separate if a firm has more than one class of equity issued. That is, *NOSH* only shows floating A-shares[66].

Common shares outstanding (“*WC05301*”) are on the other hand defined by *Datastream* as the number of shares outstanding at the firm’s year-end; the difference between total issued shares and treasury shares. That is, it includes both floating and non-floating shares, as well as other foreign-listed

[66] See Appendix for further details on datatypes ”NOSH” and ”WC05301” from *Datastream*

shares. Multiplying with common shares outstanding therefore leads to the total market capitalization, while multiplying with *NOSH* leads to the floating market capitalization of A-shares. As previously elaborated on in the methodology, this thesis argues that market capitalization should be defined as the total market capitalization, and is therefore calculated as unadjusted price (*UP*) times common shares outstanding (*WC05301*).

Book-to-market ratio

The book value of common equity (datatype *WC03501*) is needed for the construction of book-to-market (*B/M*) portfolios. As elaborated on in the methodology, the traditional book value of equity measure is adjusted in this thesis to accommodate for specific features in the Chinese market. That is done by constructing it as the book value of equity divided by the total number of shares outstanding (datatype *WC05301*), thereby obtaining the book value per A-share. Then the book value per A-share is divided by the price per A-share.

Earnings-to-price and dividend-to-price

In addition to the model inputs described above, Fama and French (1992) and Fama and French (1993) use the earnings-to-price (*E/P*) ratio and the dividend-to-price (*D/P*) ratio, since these two variables also has been shown to hold predictive power for stock prices. This paper follows their example both to see if the two variables are useful for predicting stock returns in the Chinese market and to see if the book-to-market ratio absorbs these effects in the same way as in the U.S. market. The inverse of *E/P* ratio, price divided by earnings, is obtained directly from *Datastream* (datatype *PE*). It expresses the price divided by the earnings rate per share and is therefore inverted to obtain the *E/P* ratio. The *D/P* ratio is obtained directly from *Datastream* (datatype *DY*) and expresses the dividend per share as a percentage of the share price.

5.3 Inputs for the five-factor model

The book value of equity, the outstanding number of shares, the number of shares, unadjusted price and a risk-free rate as described above are also necessary inputs for the five-factor model, that adds two more factors to the original three-factor model of Fama and French (1993). The two additional factors added in the five-factor model are operating profitability and investment.

Fama and French (2015) calculate their operating profitability variable as revenues minus cost of goods sold, selling, general and administrative expenses and interest expenses. This thesis follows their measure of operating profitability, besides excluding interest expenses because of too many missing values for *Interest Expenses – Total* (datatype *WC01075*) in *Datastream*. Hence, *Net Sales or Revenues* (datatype *WC01001*), *Cost of Goods Sold Excl Depreciation* (datatype *WC01051*) and *Selling, General & Administrative Expenses* (datatype *WC01101*) are obtained from *Datastream*. The *cost of goods sold* measure obtained from *Datastream* does not include depreciation, and *Depreciation And Depletion* (datatype *WC04049)* is therefore also downloaded and subtracted from revenues. Furthermore, the investment factor is calculated as last year's change in total assets. Thus, *Total Assets* (datatype *WC02999)* is obtained from *Datastream*. An overview of all the raw inputs from *Datastream* and what they are used for are provided in Table 1 below. The table concludes the description of raw inputs, and the next section moves on to the data cleaning process.

Table 1 - Overview Datatype

Input	**Purpose**
Total Return Index (datatype "*RI*")	Monthly stock return
Number Of Shares (datatype "*NOSH*")	Number of A-shares for floating market cap
Common Shares Outstanding (datatype "*WC05301*")	Total number of shares for total market value
Unadjusted Price (datatype "*UP*")	Total and floating market value
Common Equity (datatype "*WC03501*")	Book-to-market ratio
Dividend Yield (datatype "*DY*")	Dividend-to-price ratio
Price/Earnings Ratio (datatype "*PE*")	Earnings-to-price ratio
Net Sales or Revenues (datatype *WC01001*)	Revenues
Cost of Goods Sold Excl Depreciation (datatype *WC01051*)	Cost of goods sold
Selling, General & Administrative Expenses (datatype *WC01101*)	Selling, General & Administrative Expenses
Depreciation And Depletion (datatype *WC04049)*	Depreciation
Total Assets (datatype *WC02999)*	Total assets

5.4 Data screens

Ince and Porter (2006) point out that raw *Datastream* data is commonly used for non-U.S equity research due to its broad and deep coverage, and they assess its suitability for such research. Specifically, they document several issues with raw *Datastream* data and develop a screening procedure that they show to improve the quality of the data. They further state that the goal of their paper is to develop methods for identifying errors in *Datastream* data, so that it also can be used in markets outside of the U.S. and in particular when there is no available alternative data source. This makes their paper particularly relevant for this thesis. Schmidt et al. (2017) further build upon these technics by building pan European and country-specific value, size and momentum factors based on the screening procedure presented by Ince and Porter (2006). This section consists of a static and a dynamic screening suggested by the two papers.

Static screening

First of all, *Datastream* sometimes returns the message #ERROR for a time series request. This message means that no data is available on the datatype for the respective stock in the chosen period. Since data on all the four data types, return index (*RI*), unadjusted price (*UP*), common shares outstanding (*WC05301*) and common shareholders equity (*WC03501*) are necessary to form portfolios for the three- and four-factor model, all stocks where *Datastream* returns the #ERROR message for one of the four data types are therefore removed from the sample. The same is done for the five-factor model, where the stocks lacking any of the input necessary for the profitability and investment variables are also excluded.

Secondly, following the approach of Schmidt et al. (2017), we keep all major listings (datatype "MAJOR" = "Y"), all listings located in the domestic market (datatype *GEOGN* = *CHINA*), and all listings classified as common equity (datatype *TYPE* = *EQ*). 32 non-major firms are removed from the sample due to being non-major listings. As expected, all firms are located in China, and all listings are classified as common equity.

Thirdly, the last static screening of the sample is done on the basis of the extended name of all listings (datatype ENAME). There might be information in the extended name that indicates that it is something else than common equity. Inspired by Ince and Porter (2006), a list of names and abbreviations that indicate this is therefore built and includes the following names and phrases: *CV, CONV, CVT, FD, OPCVM, PREF, PF, PFD, PFC, PFCL, RIGHTS, RTS, UNIT, UNITS, WTS, WARR, WARRANT, WARRANTS*. In addition, it is desirable to check if some dead or delisted firms are included in the current sample, and the phrases *DEAD, DELIST, EXPD, DEL, DELEST, DELISTED* and *DEF* are therefore included and searched for. Ince and Porter (2006) do not explicitly state the names they include in their list and flag, so the list in this thesis is based on Campbell, Cowan & Salotti (2010, page 3089). The listings with extended names that contain any of these names or abbreviations are then flagged.

Furthermore, the listings that do not include a single standing "A" at the end of their name are flagged. That is, they are flagged if they do not have any of the following standalone endings to their name *"A, A"*, *A* or *"A"*. The list of flagged companies is then thoroughly examined to investigate if a listing is actually something else than common equity A-shares or if it is a dual listing of another company, or if it has just ended up on the list by coincidence. An example of the latter could be that the abbreviation is a part of its name by chance.

Naturally, listings that are deemed to be either something else than common equity or a dual listing of another company are removed from the sample, while the other group stays in the sample. The extended name list based on Campell et al. (2010) returns 32 flagged names, but upon further investigation, all of the names flagged contain the abbreviation by chance. Hence, none of them is dropped. The list of firms that do not have an "A" behind their name includes 19 firms, and all of them are dropped from the sample. This concludes the static screening process. The next section proceeds with the dynamic screening.

Dynamic screening

The dynamic screening process does not delete firms permanently from the sample. It only deletes single or multiple return observations at one point in time if the value of the variable exceeds a set value for the screening. A firm that exceeds this set value will be included again as soon as it is below this limit value. The first dynamic screening that is performed is that all listings with only one monthly price index observation are removed. It is not possible to calculate return from only one price observation. Then the dynamic screening proceeds with removing all stocks with a price of less than one unit of domestic currency In this case, less than one Yuan. This is done due to the *Datastream* practice of rounding prices to the nearest penny[67], which could cause significant differences in calculated returns when prices are small. Price index values greater than 1 000 000 are also set to missing.

When it comes to calculated returns, all returns greater than 890% are set to missing. Returns greater than 300% during one month are also set to missing values if a sharp reversal is observed the next month (>50%). Specifically, if r_t or r_{t-1} is greater than 300% and $(1 + r_t)(1 + r_{t-1}) – 1$ is less than 50%, r_t and r_{t-1} are removed from the sample. Ince and Porter (2006) use a 300% threshold, which they

67 Ince and Porter (2006)

describe as somewhat arbitrary based on trial and error. This level also seems reasonable based on the same methodology for the Chinese dataset.

At last, an important aspect of the framework this thesis operates within is how to deal with stock suspensions. A particular feature of the Chinese market is the high frequency of suspensions. However, Ince and Porter (2006) point out that it is difficult to identify trading halts in *Datastream* data, and they are therefore not able to correct for this. Hence, this thesis also has to leave the issue unattended. That concludes the discussion on dynamic screening as well as overall data treatment, and the raw data sample is reduced from 3502 firms to 3231 firms. Table 2 below shows how the average number of firms in the sample develops over the period of investigation. Now that the overall framework and inputs are described, the thesis proceeds to the analysis in the next section.

Table 2 - Average number of firms per year

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Average n firms	1309	1411	1514	1577	1804	2143	2368	2444	2504	2697	2860	3231

Section 6 – Analysis

The analysis is organized in four parts. First, the factors will be univariately sorted to have a first look at trends in the Chinese A-share returns and to see if the factors have yielded significant returns. This first look at trends will also serve to direct the focus of the rest of the analysis. Second, the underlying features of the Chinese market and the excess returns of the constructed dependent variable portfolios (*LHS*) will be investigated. These excess returns are what the risk mimicking factors (*RHS*) will have a shot at explaining in the later regressions. In the third part, the excess returns of the risk mimicking factor portfolios will be investigated in isolation to see if they look promising for predicting overall returns, as well as if they have yielded significant returns over the period. In the fourth and final part, the regressions for the different models are run, and their overall fit is compared to each other. As stated in section 5.1, the returns investigated in the analysis cover the period from July 2006 to June 2017.

6.1 Univariate sorting

As a first look at the returns in the market, the sample is univariately sorted on size, book-to-market (*B/M*) ratio, momentum, profitability and investment ranking. That is, ten equally populated portfolios are formed based on the stocks' ranking on the five respective variables. The average returns of these portfolios are displayed in **Error! Reference source not found.** The ten portfolios increase in size, *B/M* ratio, momentum, profitability and investment rank from left (Portfolio 1) to the right (portfolio 10). For example portfolio 1 (left) contains the smallest listings, the listings with the lowest *B/M* ratio, momentum, profitability, or investment rank depending on which variable they are sorted on.

The returns in the size row in Table 3 monotonically decrease in size. The portfolio containing the smallest stocks has earned an average return over the period of 2.70% per month. That is high both compared to the biggest portfolio's average return of 1.03% per month and the average 1.81% per month over the period. The trend for *B/M* portfolios is decreasing instead of increasing returns in *B/M* ratio. Furthermore, the trend of decreasing returns is rather monotonous, even though it is not as monotonous as the size trend. As for the size returns, the *B/M* portfolio return located in the lowest *B/M* portfolio (2.12%) is high compared to the return of the highest *B/M* portfolio (1.34%) and the average of 1.45%. Contrary to size, the *B/M* ratio portfolios display an unusual trend held up against previous factor literature. The value premium has gotten its name because value stocks (high *B/M* ratio) tend to outperform growth stocks (low *B/M* ratio), but the opposite trend is observed in Table 3.

The momentum trend in the Chinese A-share market is that recent losers outperform recent winners. That is, one can observe a reversal effect. However, the trend of losers outperforming winners is not as categorical as it is for the size and the *B/M* portfolios. Even though the difference between the return of the low momentum portfolio (1.44%) and the high momentum portfolio (0.92%) is high, the return on the low momentum portfolio of 1.44% is not much higher than the overall average of 1.43%. This is also illustrated by the fact that four out of nine portfolio returns increase as one moves from left (recent losers) to the right (recent winners). In fact, besides the 10^{th} portfolio containing the biggest winners over the past year, the pattern across momentum looks rather random. However, the trend is negative, which means that a reversal strategy has been more successful than a momentum strategy over the medium term during the period of investigation, in contrast to findings in other markets.

The two last factors added to introduce the five-factor model of Fama and French (2015), investment and operating profitability, also show mixed return patterns. The constructed investment portfolios display a pattern that is consistent with the suggestion of the Miller and Modigliani (*MM*) (1961) equation displayed section 3.3: firms with low investment activity earn a premium over firms with high investment activity. The sign of the conservative minus aggressive (*CMA*) portfolio is therefore as expected. This pattern is similar to the previously described *B/M* pattern; firms with low investment activity earn a high average monthly return (1.79%) compared to the return of firms with high investment activity (1.06%) and the average monthly return across the ten investment portfolios

(1.48%). However, the pattern is not monotonously decreasing from left (low investment) to the right (high investment) and does not seem consistent.

The constructed operational profitability portfolios do not display a pattern that is consistent with the proposal of the *MM* equation. The equation proposes that firms with high operational profitability should earn a premium over firms with low operational profitability, but the opposite is true in the Chinese sample. The sign of the robust minus weak (*RMW*) portfolio is therefore not as expected. However, the trend of firms with low profitability outperforming firms with high profitability is not monotonous across the portfolios. The difference between the return of the least profitable portfolio (1.88%) and the most profitable portfolio (1.36%) is high, and the return on the least profitable portfolio of 1.88% is quite a lot higher than the overall average of 1.48%. With the negative sign on the factor and the mixed pattern, the operational profitability returns are similar to the *B/M* ratio, but there are bigger differences and a clearer pattern for the *B/M* ratio.

Overall, this initial look at returns through the lens of the factors of interest shows a strong size effect with a clear pattern of returns decreasing in size. All the other factors look weaker than expected and they all display mixed patterns. Furthermore, the size and the investment effect have return patterns consistent with research in other markets, whereas *B/M*, momentum and operating profitability show return patterns in conflict with expectations. However, the unconvincing patterns of the four other factors do not mean that they should not be a part of an overall model for the Chinese equity market. This will be further discussed as the thesis proceeds, and the dependent variables in the regressions (*LHS*) are described in the next part.

Table 3 – Univariate sorted portfolios

	Factor Returns									
	Portfolio returns by Decile									
	Low/ Losser									**High/ Winner**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Size	2.70%	2.26%	2.06%	1.88%	1.87%	1.68%	1.61%	1.56%	1.44%	1.03%
B/M	2.12%	1.60%	1.66%	1.60%	1.35%	1.31%	0.96%	1.40%	1.17%	1.34%
Momentum	1.44%	1.66%	1.58%	1.32%	1.68%	1.49%	1.29%	1.62%	1.27%	0.92%
Inv	1.79%	1.82%	1.59%	1.75%	1.49%	1.53%	1.46%	1.44%	1.16%	1.06%
OP	1.88%	1.84%	1.42%	1.74%	1.42%	1.33%	1.25%	1.39%	1.16%	1.36%

6.2 Summary statistics dependent variable:

This section moves on to describe some underlying features of the Chinese data sample and the dependent variables (*LHS*) to be regressed on later. That is characteristics about the different portfolios as well as their return patterns. Fama and French (1993) use the size and *B/M* variables to form double-sorted portfolios, and Fama and French (2015) also add sorts on size-profitability and size-investment. The excess returns on portfolios sorted on size combined with momentum, profitability and investment are displayed in the Appendix. From these portfolios, in addition to the size-*B/M* sort, it can be seen that the *B/M* factor provides the most evident pattern besides size. For reasons of readability, the summary statistics for the data sample is only displayed and discussed for the size-*B/M* sort. Even though the *B/M* trend is opposite of convention and what is to be expected, it nevertheless seems to have a trend. Another advantage of focusing on the size-*B/M* sort of the dependent variable is that it allows for comparison with the observations of Fama and French (1993). For more detailed insights on the other dependent variable sorts, the reader is referred to the Appendix.

Fama and French (1993) use New York Stock Exchange breakpoints to form the 25 size and book-to-market portfolios, and therefore end up with many stocks in the smaller size quantiles[68]. This paper sets the quantile breakpoints based on the entire Chinese market. The size distribution is, therefore, more equal, with approximately the same amount of firms in each size quantile. Table 4 below displays the average number of firms in each portfolio along with other summary statistics. Even though the number of firms in each size quantile is similar, there are differences regarding average market value and average annual percentage of total market value in each portfolio. The total market value in the five smallest portfolios make up 3.3% of the total market value, ranging from 0.31% in the lowest *B/M* portfolio to 0.91% in the highest *B/M* portfolio. The five biggest portfolios on the other hand account for 66.75% of the total market value, ranging from 8.43% to 25.05%.

Regarding size, the sample of investigation is similar to the one Fama and French (1993) used, but the range of portfolio value is wider across the *B/M* ratio axis, especially in the ten smallest size portfolios. For example, the five smallest portfolios of Fama and French (1993) range from an annual average of 0.46% to 0.69% of total value, while in the Chinese sample they range from 0.31% to 0.99%. There is also a distinct pattern in the Chinese sample, with more value in the high *B/M* portfolios for small firms. This pattern is reverse for big firms, where market value increases from high *B/M* portfolios to low *B/M* portfolios. The same two patterns are also observed for the average number of firms in the 25 portfolios. The pattern from Fama and French (1993) is increasing percentage market value from high *B/M* portfolios to low *B/M* portfolios within all size quantiles, except the smallest one where there is no clear trend. Their largest portfolio is the biggest firm and lowest *B/M* portfolio with around 30% of

[68] Their sample includes NYSE, Nasdaq and Amex stocks, whereby NYSE stock are by far larger on average than stocks from the two other exchanges

total market value. This portfolio is also the largest in the Chinese market with approximately 25% of total market value.

The only exception from the market value pattern described above with increasing share of market value in *B/M* for smaller portfolios and decreasing share of market value in *B/M* for larger portfolios is observed in the largest size and highest *B/M* portfolio. This portfolio also accounts for a much larger share of the total market value than the same portfolio in Fama and French (1993), with 13.87% of the total Chinese market compared to 4.61% in the Fama and French paper. This difference mainly comes from the much higher average annual firm size in this portfolio than the other portfolios, since there are somewhat similar or fewer firms there than in portfolios with similar size but lower *B/M* ratio. This is also one of the peculiar characteristics in the Chinese market, with some giant state-owned firms. The portfolio is an even larger part of the total market capitalization when the financial firms are included, due to the huge state-owned banks. With financial firms, the large-cap and high *B/M* portfolio is 22.39% of the total market on average (See Appendix for details).

Table 4 also displays averages of annual earnings-to-price (*E/P*) and annual dividend-to-price ratios (*D/P*), as well as market leverage (*TA/M*) and book leverage (*A/B*). Controlling for size, E/P and D/P increase monotonically with higher *B/M* ratio. The two portfolios in the two smallest size quantiles and the lowest *B/M* quantile are the only two exceptions from this pattern, with much higher *E/P* than they should have had if the pattern completely held. The pattern holds completely for all 25 portfolios in the D/P matrix. This pattern is generally in line with the U.S. market. Market leverage (*TA/M*) is also in line with what is typically observed in the U.S. market; firms with high *B/M* ratio have high market leverage ratios. That is, the market deems the earnings prospects of these firms as poor, and they should, therefore, have a higher excess return due to the risk of continued poor earnings.

However, the book leverage matrix displays a slightly different pattern than what is typically observed. The book leverage ratio normally decreases in *B/M* ratio due to the higher leverage of high *B/M* firms. That only holds true for small firms in the Chinese sample, as leverage increases in *B/M* ratio for firms in the three largest quantiles. This might have important implications for the returns to be investigated. Fama and French (1992) rationalize their finding of higher returns for high *B/M* firms by the observations of Chan and Chen (1991) of more marginal firms in the portfolios with higher *B/M*. Marginal firms are categorized by high leverage and poor earnings prospects, and it will be interesting to see the effect of one of these conditions to some degree violated. Some of the distress risk could be taken away, which should mean different HML returns.

The Chinese data sample is in many ways similar to the sample Fama and French (1993) used in their initial research on the U.S. market, but there are also some aspects that are different. In particular, in the Chinese sample, the size distribution is more even regarding the number of firms due to the different procedure for dividing the sample into size portfolios, and the distinct pattern for market value in Fama and French (1993) found across *B/M* portfolios are more mixed. The latter might be a sign of a more mixed *B/M* effect in China. At last, the book leverage pattern is also different in the Chinese sample, with more portfolios with high leverage in high *B/M* portfolios. That might also have implications for the *B/M* effect, which will become clearer in the next section as the thesis moves on to the excess return patterns of the dependent variable (*LHS*).

Table 4 - Summary Statistics of portfolios sorted on Size & B/M

Sort on Size - B/M (5x5)

Average of annual number of firms in Portfolio

	Low	2	3	4	High
Small	37	51	82	105	106
2	38	71	86	93	94
3	58	90	84	77	72
4	100	97	74	60	51
Big	149	73	55	47	59

Average of annual percentage of MV in portfolio

	Low	2	3	4	High
Small	0.31	0.45	0.74	0.89	0.91
2	0.61	1.08	1.25	1.32	1.39
3	1.44	2.17	1.9	1.73	1.62
4	4.31	3.84	2.89	2.37	2.03
Big	25.05	9.24	8.43	10.16	13.87

Average of Average Annual firm Size in Total MV

	Low	2	3	4	High
Small	2655	2805	2896	2708	2770
2	5241	4951	4573	4501	4685
3	8047	7629	7185	7055	7092
4	13456	12500	12236	12407	12869
Big	51082	38514	45275	68676	75257

Average of annual B/M ratios for portfolio

	Low	2	3	4	High
Small	0.05	0.13	0.21	0.33	0.58
2	0.06	0.13	0.21	0.33	0.59
3	0.06	0.13	0.21	0.33	0.60
4	0.06	0.12	0.21	0.33	0.62
Big	0.05	0.12	0.21	0.33	0.62

Average of annual E/P ratio in % for portfolios

	Low	2	3	4	High
Small	2.51%	1.49%	1.73%	1.92%	2.59%
2	2.05%	1.97%	2.23%	2.53%	3.13%
3	2.25%	2.31%	2.78%	3.11%	3.77%
4	2.40%	2.79%	3.30%	3.80%	4.74%
Big	2.86%	3.83%	4.35%	4.96%	6.12%

Average of annual D/P ratios in % for portfolio

	Low	2	3	4	High
Small	0.03%	0.16%	0.26%	0.34%	0.61%
2	0.24%	0.40%	0.51%	0.57%	0.68%
3	0.44%	0.53%	0.60%	0.70%	0.96%
4	0.52%	0.68%	0.89%	0.96%	1.41%
Big	0.64%	1.01%	1.16%	1.33%	1.88%

Leverage Ratio TA/Float_MV

	Low	2	3	4	High
Small	0.41	0.42	0.50	0.75	1.21
2	0.25	0.37	0.50	0.86	1.37
3	0.21	0.34	0.57	0.84	1.55
4	0.18	0.35	0.58	0.92	1.77
Big	0.18	0.41	0.70	1.08	2.06

Leverage Ratio TA/EQ

	Low	2	3	4	High
Small	16.32	4.67	3.22	2.55	2.97
2	4.12	2.58	2.50	2.56	2.79
3	2.43	2.27	2.81	2.60	2.85
4	2.17	2.26	2.47	2.85	2.93
Big	2.28	2.60	2.78	2.95	3.16

6.3 Excess returns on dependent variable portfolios:

The excess returns on the 25 size-*B/M* sorted dependent variable portfolios in Table 5 illustrate the range of excess returns that the factors under investigation should explain. From the 5x5 matrix of monthly excess returns, it can be observed that the portfolio returns range from 0.52% to 2.47%. The matrix displays a clear size pattern within the *B/M* quantiles; controlling for *B/M*, average excess returns decrease monotonically as the size of the firms in the portfolios increase. Small firms seem to earn higher average excess returns than big firms. That is consistent with the initial findings of Fama and French (1993) and most factor research done in the Chinese market. The t-value matrix also illustrates the difference between small and big firms. It shows that t-values for the five smallest portfolios range between 2.06 and 2.59, while t-values for the five biggest portfolios range between 0.47 and 1.74. Thus all excess returns in the five smallest portfolios are significantly different from zero on a 5%-level, while none of the excess returns in the five biggest portfolios is significantly different from zero on the same 5%-level.

In contrast to the monotonically decreasing excess returns from small to big firms, the *B/M* pattern is more mixed. The excess returns display a mixed pattern for *B/M* within each size quantile, but the pattern seems to be the same as observed for the univariately sorted portfolios; lower excess returns for portfolios with higher *B/M* ratios. That would stand in strong contrast to the well-documented value premium in other markets. However, the book-to-market pattern is weaker than the previously observed size pattern, with excess returns also increasing in *B/M* ratio on several occasions. Furthermore, one out of five low *B/M* portfolios is not being significantly different from zero on the 5%-level and one out of five high *B/M* portfolios is significantly different from zero. Nevertheless, the pattern of decreasing excess returns in *B/M* ratio looks somewhat clear, even though it is not completely monotonous.

Double sorted size-momentum, size-profitability and size-investment excess return matrices can be found in the Appendix. Similar to the *B/M* pattern, all of the other patterns display some trend, but these trends are mixed. The strongest trend can be observed for the size-momentum sort, where recent losers outperform recent winners as the univariately sorted portfolios indicated. However, the pattern is even more mixed for momentum than for *B/M*. The size-profitability and size-investment sorted portfolios confirm the patterns observed for the univariately sorted profitability and investment portfolios. That is low profitability and low investment portfolios earning a premium relative to the high profitability and high investment portfolios. However, these patterns are even more mixed than the previously described *B/M* and momentum patterns. By holding size constant and moving across the profitability and investment sorting, the patterns at times seem rather random. The thesis now moves on to estimating the different premia in the next part, which will be interesting considering the unusual patterns some of the factors show.

Table 5 - Multivariate sort on Size & B/M

Sort on Size - B/M (5x5)

	Excess Returns				
	Low	**2**	**3**	**4**	**High**
Small	2.46%	2.47%	2.42%	2.16%	2.23%
2	2.16%	2.27%	1.85%	1.64%	1.67%
3	2.00%	2.00%	1.59%	1.38%	1.27%
4	2.18%	1.36%	1.21%	1.22%	1.17%
Big	1.57%	1.26%	0.81%	0.52%	0.95%

	Standard Deviation				
	Low	**2**	**3**	**4**	**High**
Small	3.15%	3.53%	3.38%	3.48%	3.41%
2	3.47%	3.51%	3.30%	3.33%	3.42%
3	3.76%	3.12%	3.28%	3.31%	3.74%
4	3.58%	3.39%	3.58%	3.75%	3.94%
Big	2.99%	3.84%	3.65%	3.65%	3.70%

	t-values				
	Low	**2**	**3**	**4**	**High**
Small	2.59	2.32	2.37	2.06	2.16
2	2.06	2.14	1.85	1.63	1.62
3	1.76	2.12	1.61	1.38	1.12
4	2.01	1.33	1.12	1.08	0.98
Big	1.74	1.09	0.74	0.47	0.85

6.4 Explanatory returns (*RHS*)

Next up is the independent variable (*RHS*) in the later regressions, which will have a shot at explaining the dependent variable (*LHS*) return patterns described in the previous section. The independent variables in the time-series approach are the monthly returns of the risk factor mimicking portfolios. That is, the average of the returns of these portfolios are the average premium per unit of risk (exposure) in the APT framework. Table 6 displays the means, standard deviations and t-statistics of the excess market return (*RM-RF*), as well as small-minus-big (*SMB*), high-minus-low (*HML*), winner-minus-loser (*WML*), robust-minus-weak (*RMW*), and conservative-minus-aggressive (*CMA*) portfolio returns. The means of these factors are respectively the estimated market, small size, value, momentum, profitability and investment premia over the period.

The market premium is economically large with a monthly mean of 0.97%, which translates into 12.28% on an annualized basis. It is however not statistically significant due to the high volatility of the market factor of 8.9% per month or approximately 31% annually. These observations are consistent with much of the factor research on the Chinese A-share market, as well as Fama and French (1993). As indicated by the clear pattern in both univariately and double sorted portfolio excess returns, the *SMB* factor premium is both economically large and statistically significant on a 5%-level. On average it earns 0.86% monthly or 10.89% annually. This is also consistent with most of the Chinese factor literature, while Fama and French (1993) only found an insignificant size premium.

As the pattern from the ten univariately sorted portfolios and the 25 dependent excess return portfolios indicated, the value premium is indeed negative. That is, growth firms (low *B/M*) earn an average premium over value firms (high *B/M*) of 0.54% per month (6.26% annually) from July 2006 to June 2017. However, the premium is not statistically significant, even though its monthly standard deviation of 4.46% or approximately 15% annually is low compared to *SMB* and the market factor. In the Chinese factor literature, the lack of a value premium has been found among others by Hu et al. (2018). On the other hand, it is inconsistent with Fama and French (2012) and Asness, Moskowitz and Pedersen (2013) finding value premia across numerous global markets. Growth firms earning a premium over value firms is even more uncommon, but it is difficult to draw any solid conclusions from this as long as the premium is insignificant.

As the value premium, momentum has the opposite sign of what one would expect. The momentum strategy of buying recent winners and selling recent losers has been found successful over the medium term in many different markets. Nevertheless, over the period from July 2006 to June 2017, this strategy has earned negative average returns in the Chinese A-share market. The momentum strategy tested here is constructed following Carhart (1997), which means that it ranks stocks based on their past 12 months return minus the last month and rebalances every month. The negative returns from this strategy are not just economically large, but also statistically significant on the 5%-level. That is, in the period from July 2006 to June 2017 a reversal strategy has earned a statistically significant monthly return of 0.71% (8.90% annually) in the Chinese A-share market.

At last, the profitability and investment factors have earned very low returns in the Chinese A-share market over the period. The *RMW* portfolio has earned a monthly return of -0.28%, or -3.35% annually, while the *CMA* portfolio has earned a 0.19% monthly return or 2.28% annually. Recall from the univariately and double-sorted portfolios that the trends for both profitability and investment look weak. The estimated *RMW* and *CMA* premia further support this picture. The factors have certainly not yielded high returns over the period investigated, and will therefore not be further investigated in this section. However, they will naturally be a part of the five-factor model tests.

Table 6 - Factor premia

	RM	SMB	HML	WML	RMW	CMA
Mean	0.97%	0.86%	-0.53%	-0.70%	-0.28%	0.19%
SD	8.90%	4.67%	4.27%	3.67%	2.91%	1.83%
t-value	1.25	2.12	-1.43	-2.18	-1.08	1.18

To investigate the performance of value and momentum more detailed, Table 7 shows the *HML* and *WML* factors broken down to *HML* for small and big firms, and *WML* for small and big firms. The value premium is slightly less negative for small firms than for big firms. The differences in leverage from the expected characteristics of high *B/M* firms mentioned in the descriptive statistics part therefore turn out to be of low importance. If those differences were of importance, the return on *HML* for small firms should be much bigger than the return on *HML* for big firms. Furthermore, the *WML* premium is less negative for small firms than for big firms.

Table 7 - Closer look at factor premia

	RM	SMB	HML	HMLS	HMLB	HML S–B	WML	WMLS	WMLB	WMLS-B
Mean	0.97%	0.86%	-0.54%	-0.49%	-0.59%	0.10%	-0.71%	-1.18%	-0.25%	-0.93%
SD	8.90%	4.66%	4.34%	3.26%	6.08%		3.66%	3.03%	5.09%	
t-value	1.252062	2.132028	-1.42385	-1.71112	-1.11305		-2.23543	-4.47601	-0.55425	

Furthermore, from the correlations in Table 8 it can be seen that the Chinese market factor is weakly positively correlated with *SMB* and *HML*, 6.0% and 13.9% respectively. The *SMB* and *HML* factors have a stronger negative correlation of -33.9%. Fama and French (1993) had quite strong positive correlations between the market factor and *SMB*, a somewhat equally strong negative correlation between the market factor and *HML*, and almost no correlation between *HML* and *SMB*. The lack of correlation between *HML* and *SMB* was highlighted as one of the reasons for the wide range of cross-sectional returns they covered by using these two factors. Another thing that was important was the strong opposite correlation with the market *HML* and *SMB* had. The correlation structure between U.S. market sample Fama and French (1993) used and the one used in this thesis is therefore quite different.

The momentum factor is almost not correlated with the market at all, and it is strongly negatively correlated with the *HML* factor. It is also negatively correlated with the size factor, even though this negative correlation is very weak (-0.039). The *RMW* and *CMA* factors are both negatively correlated with the market, similar to the observations of Fama and French (2015). It is also interesting to note that there is a very strong negative correlation between *RMW* and *SMB* of -0.75, as well as a positive correlation between *HML* and *RMW*. Especially the latter will be further discussed in section 8, since the value and profitability factor premia have been shown to be negatively correlated with each other in other markets[69].

Table 8 - Factor correlations

Correlations						
	RM	**SMB**	**HML**	**WML**	**RMW**	**CMA**
RM	1					
SMB	0.06	1				
HML	0.139	-0.334	1			
WML	-0.008	-0.039	-0.411	1		
RMW	-0.274	-0.748	0.243	-0.095	1	
CMA	-0.201	0.241	0.375	-0.161	-0.274	1

This investigation of the independent variable (*RHS*) shows a strong but insignificant market premium, a strong and significant size premium, and a lack of value, profitability and investment premia. All of these findings are expected based on the observations previously made for univariately and double-sorted portfolios in previous sections, but the lack of value, profitability and investment premia are still unexpected based on research in other markets. Also, there is not only a lack of a medium-term momentum effect, but a statistically significant reversal effect is observed. That concludes the introduction to the dependent and the independent variables in the regressions, and the thesis now moves on to the regressions for different factor models.

6.5 Multifactor regressions

After having outlined the dependent and independent variables, the four different models this thesis set out to test will now be formally tested. First up is the *CAPM*, followed by the three-factor model, the four-factor model and the five-factor model. The methodical set-up is described in section four and will be executed in this and following sections.

CAPM

Table 9 shows the regression output for the standard *CAPM* regression. The R-squared values for the *CAPM* regression range from a moderate 0.57 for the smallest and lowest *B/M* quantile to a high 0.91 for the biggest and 3rd highest *B/M* portfolio. Only three portfolios are close to 0.9. The main part of

[69] Novy-Marx (2013)

the R-squared values ranges between 0.6 and 0.8, with 17 out of 25 portfolios covered by this range. This means that much of the overall variation is still left unexplained, especially within the two smallest size quantiles where all the R-squared values are between 0.55 and 0.75. Many of the alphas of the regression are economically large, ranging between -0.48% and 1.55%. As first evidence of a not completely specified model, there are nine positive and statistically significant, and the average alpha is 0.711%. The trend for alphas is very clear when size isolated from *B/M* ratio is investigated, with alphas monotonically decreasing from smaller to bigger firms. This is similar to what the pattern of R-squared values shows. There is also a pattern of smaller R-squared values and larger alphas in portfolios with low *B/M* ratios. The conditions are therefore good for both size and *B/M* ratio to add a significant amount of explanatory power to the model.

The market ß's are all close to one, with 24 out of 25 portfolio ß's being less than 0.1 from one. The market ß's seem to increase in *B/M* ratio, which means that the portfolios with high *B/M* ratios are more dependent on the market than the low *B/M* ratio portfolios. That also seems to be the picture in the R-squared and alpha matrix, where the market explains less in the low *B/M* portfolios. That means that return differences between high and low *B/M* portfolios can be driven by different exposures to the market. However, the expectation would be to observe that low *B/M* portfolios are more exposed to the market, which could help explain the negative value premium. As for size, there is no detectable pattern for the ßs. Differences in excess returns between small and large firms are therefore not likely to be driven by different exposure to the market.

Table 9 - CAPM regression output

CAPM

Sort on Size - B/M (5x5)

	Market ß				
	Low	**2**	**3**	**4**	**High**
Small	0.94	1.04	1.06	1.06	1.05
2	1.01	1.09	1.06	1.07	1.09
3	1.05	1.07	1.08	1.09	1.10
4	1.01	1.03	1.06	1.12	1.10
Big	0.96	1.05	1.06	1.03	1.01

	t(ß)				
	Low	**2**	**3**	**4**	**High**
Small	10.874	10.116	12.483	11.343	11.958
2	11.055	11.994	12.752	12.875	13.938
3	12.579	13.424	13.837	14.069	14.479
4	13.878	13.963	15.484	17.244	19.393
Big	21.769	22.048	25.625	24.990	17.897

	R^2				
	Low	**2**	**3**	**4**	**High**
Small	0.57	0.61	0.67	0.66	0.67
2	0.62	0.69	0.70	0.71	0.74
3	0.70	0.71	0.74	0.77	0.76
4	0.68	0.73	0.79	0.83	0.83
Big	0.82	0.87	0.91	0.87	0.79

	α				
	Low	**2**	**3**	**4**	**High**
Small	1.55%	1.46%	1.40%	1.13%	1.21%
2	1.18%	1.21%	0.82%	0.60%	0.61%
3	0.98%	0.96%	0.55%	0.32%	0.20%
4	1.19%	0.37%	0.19%	0.14%	0.10%
Big	0.64%	0.24%	-0.22%	-0.48%	-0.02%

	t(α)				
	Low	**2**	**3**	**4**	**High**
Small	2.548	2.359	2.534	1.993	2.242
2	2.052	2.291	1.635	1.188	1.286
3	1.954	1.946	1.197	0.735	0.453
4	2.297	0.808	0.468	0.375	0.284
Big	1.866	0.809	-0.842	-1.576	-0.062

Two-factor model

Table 10 displays the output of the excess portfolio returns regressed on *SMB* and *HML*. As in Fama and French (1993), *SMB* and *HML* seem to capture much of the cross-sectional variations in returns. Most R-squared values are of a certain size, with 19 out of 25 above or slightly below 0.2. The highest R-squared is 0.37, and the most explanatory power seems to lay in the 15-20 smallest portfolios. The R-squared numbers are low compared to Fama and French (1993), who for example report eight values above 0.5. However, the correlation structure between the two samples are also quite different with stronger correlations in the U.S. sample and almost no correlation in the Chinese sample, so that could also play a role. Fama and French (1993) observe several significant intercepts in their regression, while the regressions in Table 10 do not have any significant intercepts. Even though the intercepts are not significant, they still have high values (average of 0.936% compared to *CAPM* average of 0.711%) and adding the market should be able to bring them further down.

Three-factor model

Table 11 shows the regression output of the three-factor model. As alpha and R-squared patterns from the *CAPM* regression suggest, the size and *B/M* factors add a significant amount of explanatory power to the model. R-squared values rise from an average of 74% to an average of 92%, and only one of the 25 alphas is significantly different from zero. The average alpha is of 0.245% is considerably lower than for *CAPM* and the combined *HML* and *SMB*. More importantly, there is no clear pattern in the R-squared and alpha matrices anymore. The R-squared and alpha values for high *B/M* portfolios seem to be respectively higher and lower than for low *B/M* portfolios, so that might be something the additional factors added later can be useful for correcting. The market ß's are closer to one than observed in the *CAPM* regression, with all of them less than 0.1 different from one. In fact, only eight of the 25 ß's are more than 0.05 away from one.

The two other coefficients in the regressions, *SMB* (s) and *HML* (h) also provide some valuable insights. First, the factor exposure towards *SMB* decreases in Size and the factor exposure towards *HML* increases in *B/M* ratio. This makes sense all the time the portfolios are sorted on Size and *B/M* ratio. Secondly, the *SMB* factor loading is relatively large in the 20 smallest portfolios, ranging between 0.66 and 1.39, and close to zero for the five largest portfolios. These *SMB* factor loadings are overall similar to the findings of Fama and French (1993), but they are almost constantly higher. Combined with the high average *SMB* premium, the small firm effect seems to hold much explanatory power in the Chinese A-share market. Thirdly, the *HML* factor loading is small not only compared to the *SMB* factor loading but also when compared to the results of Fama and French (1993). Only six *HML* coefficients have a value higher than 0.4, compared to 13 in Fama and French (1993). Combined with the previously discussed large *SMB* premium and factor exposure, the low exposure towards the *HML* factor is interesting due to what it might imply regarding the importance of the factor.

Table 10 - Two-factor regression output

Two Factor Regression

Sort on Size - B/M (5x5)

Factor loading SMB (s)

	Low	2	3	4	High
Small	1.34	1.60	1.49	1.61	1.57
2	1.45	1.34	1.37	1.41	1.40
3	1.15	1.19	1.21	1.23	1.27
4	0.99	1.04	0.89	0.91	0.90
Big	0.25	0.42	0.22	0.23	0.01

t(s)

	Low	2	3	4	High
Small	5.905	7.177	6.266	6.926	6.871
2	6.447	5.695	6.036	5.976	5.775
3	5.267	5.425	5.107	4.926	5.289
4	4.315	4.580	3.658	3.379	3.908
Big	1.065	1.650	0.870	1.026	0.046

Factor loading HML (h)

	Low	2	3	4	High
Small	0.04	0.18	0.32	0.50	0.65
2	0.17	0.03	0.23	0.48	0.84
3	-0.03	0.03	0.23	0.50	0.89
4	-0.30	0.02	0.19	0.54	0.98
Big	-0.36	0.15	0.37	0.62	1.10

t(h)

	Low	2	3	4	High
Small	0.225	0.686	1.339	2.098	2.920
2	0.749	0.138	1.017	2.055	3.700
3	-0.100	0.101	0.947	1.983	3.850
4	-1.214	0.092	0.762	2.000	4.357
Big	-1.625	0.611	1.385	2.697	5.131

R^2

	Low	2	3	4	High
Small	0.30	0.37	0.32	0.36	0.36
2	0.32	0.27	0.29	0.29	0.30
3	0.22	0.23	0.22	0.23	0.27
4	0.22	0.19	0.13	0.13	0.19
Big	0.04	0.02	0.01	0.05	0.21

α

	Low	2	3	4	High
Small	1.32%	1.18%	1.30%	1.04%	1.22%
2	0.99%	1.13%	0.79%	0.68%	0.91%
3	0.99%	0.98%	0.67%	0.58%	0.64%
4	1.15%	0.48%	0.54%	0.73%	0.92%
Big	1.16%	0.98%	0.83%	0.65%	1.53%

t(α)

	Low	2	3	4	High
Small	1.59311	1.36881	1.51089	1.23367	1.48636
2	1.19089	1.26767	0.9291	0.78665	1.06412
3	1.12887	1.12234	0.76885	0.66036	0.74534
4	1.31176	0.55658	0.60928	0.78977	1.06749
Big	1.3813	1.08815	0.94406	0.75553	1.86092

Table 11 - Three-factor regression output

Three Factor Model Regrssion

Sort on Size - B/M (5x5)

Market ß	Low	2	3	4	High
Small	0.92	1.01	1.02	1.01	0.99
2	0.98	1.07	1.03	1.03	1.03
3	1.05	1.06	1.06	1.05	1.03
4	1.03	1.03	1.05	1.09	1.04
Big	1.00	1.06	1.06	1.01	0.97

t(ß)	Low	2	3	4	High
Small	20.39	22.92	30.27	29.28	31.55
2	22.44	26.24	34.73	30.86	33.59
3	22.94	27.48	28.63	31.09	29.12
4	30.51	24.74	23.26	26.39	29.27
Big	34.24	24.45	26.28	23.21	28.59

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.14	1.39	1.27	1.39	1.36
2	1.25	1.11	1.16	1.19	1.18
3	0.93	0.96	0.98	1.01	1.05
4	0.77	0.82	0.66	0.68	0.68
Big	0.04	0.19	-0.01	0.02	-0.20

t(s)	Low	2.00	3.00	4.00	High
Small	12.20	14.10	17.69	17.40	18.24
2	13.36	11.10	15.88	14.73	16.80
3	9.57	12.80	11.06	12.12	11.19
4	9.82	8.42	6.54	5.82	7.83
Big	0.56	1.67	-0.10	0.18	-2.65

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	-0.27	-0.16	-0.03	0.16	0.32
2	-0.16	-0.33	-0.11	0.13	0.49
3	-0.38	-0.33	-0.12	0.14	0.54
4	-0.65	-0.33	-0.17	0.17	0.63
Big	-0.70	-0.21	0.01	0.28	0.77

t(h)	Low	2.00	3.00	4.00	High
Small	-2.65	-1.80	-0.47	2.45	5.72
2	-2.33	-4.34	-1.74	2.05	8.54
3	-4.51	-5.03	-1.96	1.74	8.34
4	-8.77	-4.25	-1.80	1.56	9.93
Big	-13.35	-2.39	0.16	3.20	11.38

R^2	Low	2.00	3.00	4.00	High
Small	0.84	0.93	0.94	0.95	0.95
2	0.90	0.94	0.95	0.94	0.95
3	0.91	0.93	0.93	0.94	0.94
4	0.91	0.91	0.89	0.90	0.93
Big	0.93	0.89	0.90	0.88	0.93

α	Low	2	3	4	High
Small	0.43%	0.20%	0.32%	0.06%	0.26%
2	0.04%	0.09%	-0.21%	-0.32%	-0.08%
3	-0.02%	-0.04%	-0.35%	-0.44%	-0.36%
4	0.15%	-0.52%	-0.47%	-0.32%	-0.09%
Big	0.19%	-0.05%	-0.20%	-0.32%	0.60%

t(α)	Low	2	3	4	High
Small	1.192	0.727	1.262	0.260	1.138
2	0.137	0.338	-0.903	-1.292	-0.398
3	-0.069	-0.152	-1.392	-1.763	-1.460
4	0.556	-1.847	-1.500	-1.110	-0.359
Big	0.882	-0.151	-0.723	-1.013	2.407

Three-factor model for alternative sorts:

Beyond the market, *SMB* and *HML* factors, this thesis also investigates models that include a momentum factor, a profitability factor and an investment factor. As mentioned earlier, there still seems to be a weak trend across *B/M* portfolios. To check the ability of the factors to improve the model further, the alphas of the three-factor regressions performed on size-momentum, size-profitability and size-investment sorted portfolios are displayed in Table 12. It can be seen that the three-factor model has more trouble with the size-momentum sorting, where the regressions yield seven significant intercepts. Particularly the smallest portfolios and the portfolios with high momentum rank are a problem. The average alpha of 0.35% is also higher than for the size-*B/M* sort. Furthermore, it seems to be a pattern of higher alphas for lower momentum portfolios. This bodes well for the potential of increased model performance by adding the momentum factor that has been shown to have higher returns in the low momentum portfolios.

No alphas in the size-profitability and size-investment sorts are significantly different from zero, which means that the three-factor model is a good description of the market for these sorts. At last, it does not seem to be any trend across profitability portfolios, while it is a weak trend with higher returns for low investment portfolios across investment portfolios. The previously observed return patterns for both the investment and profitability factors have also been weak. In contrast to the momentum factor, it does therefore not seem very promising for the addition of the profitability and the investment factors. Nevertheless, this thesis will stick to what it set out to do and test the five-factor model. First, the four-factor model is tested in the next section.

Table 12 - Three-factor model regression alternative sorts

Three Factor Model Regrssion

Sort on Size - Momentum (5x5)

	α					t(α)				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	1.15%	0.52%	0.56%	-0.11%	-1.07%	3.787	2.164	2.391	-0.440	-2.949
2	-0.06%	0.37%	0.15%	-0.41%	-1.12%	-0.252	1.584	0.625	-1.597	-4.306
3	-0.02%	0.09%	-0.04%	-0.43%	-0.85%	-0.071	0.356	-0.142	-1.649	-2.680
4	-0.22%	0.01%	-0.01%	-0.27%	-0.80%	-0.674	0.030	-0.020	-0.878	-2.532
Big	-0.04%	0.13%	0.11%	0.07%	-0.11%	-0.112	0.335	0.379	0.255	-0.303

Sort on Size - OP

	α					t(α)				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	0.28%	0.13%	0.13%	0.09%	0.31%	1.088	0.505	0.501	0.346	0.873
2	-0.21%	-0.14%	-0.19%	-0.11%	-0.28%	-0.866	-0.614	-0.776	-0.474	-1.129
3	-0.48%	-0.33%	-0.26%	-0.11%	-0.13%	-1.793	-1.167	-1.018	-0.456	-0.470
4	-0.29%	-0.22%	-0.32%	-0.27%	-0.29%	-0.979	-0.784	-1.092	-1.065	-0.994
Big	-0.03%	-0.10%	-0.22%	0.10%	0.27%	-0.075	-0.271	-0.810	0.394	1.330

Sort on Size - Inv

	α					t(α)				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	0.23%	0.17%	0.30%	-0.14%	0.25%	0.950	0.693	1.116	-0.511	0.739
2	-0.07%	-0.13%	-0.13%	-0.26%	-0.30%	-0.295	-0.556	-0.558	-1.040	-1.184
3	-0.34%	-0.34%	-0.09%	-0.23%	-0.61%	-1.414	-1.297	-0.370	-0.963	-2.270
4	0.14%	-0.26%	-0.21%	-0.32%	-0.42%	0.488	-0.858	-0.762	-1.154	-1.518
Big	0.25%	0.12%	0.32%	0.13%	-0.17%	0.647	0.373	1.225	0.556	-0.671

Four-factor model

Table 13 shows the regression output for the four-factor model sorted on size-*B/M*, with the momentum coefficient on the *WML* factor, R-squared values and alphas. The other three coefficients and their t-values can be found in the Appendix. Six of the coefficients on the *WML* factor is significantly different from zero, and R-squared values are similar to the three-factor model with an average of 0.92. The alphas are very similar to the three-factor model with one of 25 intercepts significantly different from zero, but the four-factor model is slightly better with an average alpha of 0.237% compared to 0.245% for the three-factor model.

Table 13 - Four-factor model regression

Four Factor Model

Sort on Size - B/M (5x5)

Factor loading WML (w)	Low	2	3	4	High
Small	-0.32	0.06	0.07	0.07	0.08
2	0.12	0.18	0.17	0.14	0.03
3	0.18	0.14	0.07	0.11	0.13
4	0.12	0.22	0.13	0.14	-0.03
Big	0.14	0.03	-0.10	0.01	0.09

t(w)	Low	2	3	4	High
Small	-2.931	0.584	1.134	1.012	1.283
2	1.417	2.418	2.429	1.899	0.484
3	1.944	1.766	0.789	1.348	1.652
4	1.191	2.252	1.355	1.292	-0.307
Big	2.479	0.214	-1.070	0.109	1.011

R^2	Low	2	3	4	High
Small	0.85	0.93	0.94	0.95	0.95
2	0.91	0.94	0.95	0.94	0.95
3	0.91	0.93	0.93	0.94	0.94
4	0.92	0.92	0.89	0.91	0.93
Big	0.93	0.89	0.90	0.87	0.93

α	Low	2	3	4	High
Small	0.17%	0.25%	0.37%	0.12%	0.33%
2	0.01%	0.23%	-0.07%	-0.20%	-0.06%
3	0.13%	0.08%	-0.34%	-0.35%	-0.25%
4	0.24%	-0.35%	-0.35%	-0.21%	-0.11%
Big	0.25%	-0.02%	-0.28%	-0.37%	0.67%

t(α)	Low	2	3	4	High
Small	0.487	0.887	1.538	0.490	1.402
2	0.045	0.852	-0.306	-0.808	-0.278
3	0.414	0.293	-1.280	-1.408	-1.054
4	0.853	-1.277	-1.151	-0.768	-0.451
Big	1.136	-0.083	-1.011	-1.211	2.762

Table 14 shows alpha values of the four-factor model portfolios sorted on size-momentum, size-profitability and size-investment. The alphas of the size-momentum sorting are more well behaved than in the three-factor model, but there are still four significant alphas and the average alpha of 0.34% is only slightly lower than for the three-factor model. Even though adding momentum slightly reduces the problems the three-factor model has with small and high momentum portfolios, there is still a tendency of higher alphas in the low momentum portfolios. For the size-profitability and size-investment sorted portfolios, it is difficult to unveil a pattern. As for the three-factor model, it is also here zero significant intercepts, and the average alpha is even lower with 0.15% for size-profitability and 0.19% for size-investment. Overall, the four-factor model seems to describe the equity returns slightly better than the three-factor model. Next, the five-factor model will be tested.

Table 14 - Four-factor model regression for alternative sorts

Four Factor Model Regrssion

Sort on Size - Momentum (5x5)

	α					t(α)				
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	0.85%	0.43%	0.55%	0.15%	-0.73%	**Small**	3.234	1.776	2.369	0.671	-1.871
2	-0.23%	0.32%	0.29%	-0.20%	-0.73%	**2**	-0.931	1.307	1.231	-0.790	-3.156
3	-0.26%	-0.03%	0.12%	-0.24%	-0.40%	**3**	-1.103	-0.123	0.450	-0.971	-1.458
4	-0.52%	-0.07%	0.07%	-0.04%	-0.29%	**4**	-1.690	-0.253	0.259	-0.128	-1.121
Big	-0.51%	-0.25%	0.21%	0.29%	0.61%	**Big**	-1.847	-0.718	0.736	1.009	2.323

Sort on Size - OP

	α					t(α)					
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	0.24%	0.27%	0.25%	0.10%	0.22%	**Small**	0.928	1.062	1.003	0.350	0.641
2	-0.12%	0.01%	-0.10%	-0.03%	-0.28%	**2**	-0.470	0.034	-0.389	-0.111	-1.074
3	-0.33%	-0.18%	-0.16%	-0.02%	-0.16%	**3**	-1.325	-0.644	-0.640	-0.080	-0.593
4	-0.20%	-0.02%	-0.14%	-0.16%	-0.28%	**4**	-0.676	-0.074	-0.465	-0.651	-1.005
Big	-0.01%	0.06%	-0.02%	0.29%	0.17%	**Big**	-0.017	0.167	-0.082	1.087	0.834

Sort on Size - Inv

	α					t(α)					
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	0.20%	0.28%	0.35%	-0.15%	0.22%	**Small**	0.804	1.230	1.299	-0.555	0.639
2	0.01%	-0.03%	0.01%	-0.18%	-0.27%	**2**	0.032	-0.110	0.056	-0.707	-1.084
3	-0.23%	-0.19%	0.00%	-0.22%	-0.51%	**3**	-0.944	-0.762	0.007	-0.907	-1.922
4	0.26%	-0.09%	-0.04%	-0.30%	-0.31%	**4**	0.899	-0.290	-0.150	-1.071	-1.133
Big	0.23%	0.28%	0.33%	0.09%	-0.10%	**Big**	0.595	0.880	1.275	0.377	-0.404

Five-factor model

Table 15 shows the regression output for the five-factor model sorted on size-*B/M*. It displays the profitability and investment factor loadings, as well as R-squared values and alphas. The three other coefficients can be found in the Appendix, as they are similar to the factor loadings in the previous models. It is worth noting that the market ß's are very close to one with only one out of 25 market ß's deviating with more than 0.05 from one. R-squared values are similar to the four-factor model with an average of 0.92, the alphas are similar to the three- and four-factor model with one of 25 intercepts significantly different from zero, and the absolute average alpha of 0.222% is lower than for both the three- and four-factor model. Also, most of the profitability loadings and several of the investment loadings are significantly different from zero. At last, the alpha pattern seems to be rather random compared to the three-factor model, but judging by the R-squared values, there is still more explanatory power in the high *B/M* portfolios than in the low *B/M* portfolios. Taking all of this into consideration, the profitability and investment factors may add some additional explanatory power beyond the three-factor model.

Table 15 - Five-factor model regression

Five Factor Model Regrssion

Sort on Size - B/M (5x5)

Factor loading RMW (r)	Low	2	3	4	High	t(r)	Low	2	3	4	High
Small	0.08	-0.36	-0.30	-0.35	-0.33	Small	0.420	-2.318	-2.136	-2.227	-2.051
2	-0.28	-0.46	-0.51	-0.39	-0.11	2	-1.589	-3.826	-3.502	-2.465	-0.794
3	-0.81	-0.42	-0.44	-0.37	-0.35	3	-5.348	-2.915	-2.698	-2.319	-1.941
4	-0.47	-0.46	-0.52	-0.41	-0.23	4	-2.541	-2.300	-2.824	-2.226	-1.295
Big	-0.21	-0.37	-0.34	-0.29	-0.35	Big	-1.320	-2.105	-1.665	-1.532	-2.115

Factor loading CMA (c)	Low	2	3	4	High	t(c)	Low	2	3	4	High
Small	0.45	0.14	0.03	-0.18	-0.16	Small	1.629	0.579	0.164	-1.000	-0.913
2	-0.08	-0.38	-0.39	-0.40	-0.14	2	-0.290	-2.065	-2.463	-2.310	-0.805
3	-0.42	-0.29	-0.45	-0.24	-0.25	3	-1.720	-1.454	-2.391	-1.330	-1.214
4	-0.16	-0.26	-0.27	-0.26	0.00	4	-0.822	-1.285	-1.184	-1.225	-0.013
Big	-0.28	-0.49	-0.36	0.25	-0.21	Big	-1.544	-2.064	-1.535	1.013	-0.978

R^2	Low	2	3	4	High
Small	0.85	0.93	0.94	0.95	0.95
2	0.91	0.94	0.95	0.94	0.95
3	0.91	0.93	0.93	0.94	0.94
4	0.92	0.92	0.89	0.91	0.93
Big	0.93	0.89	0.90	0.87	0.93

Size	Low	2	3	4	High		Low	2	3	4	High
	α						t(α)				
Small	0.33%	0.26%	0.36%	0.19%	0.36%	Small	0.894	0.982	1.479	0.788	1.515
2	0.10%	0.12%	-0.02%	-0.13%	-0.02%	2	0.319	0.595	-0.104	-0.561	-0.093
3	0.27%	0.12%	-0.14%	-0.33%	-0.22%	3	1.018	0.440	-0.587	-1.311	-0.884
4	0.24%	-0.38%	-0.28%	-0.17%	-0.01%	4	0.898	-1.438	-0.917	-0.600	-0.055
Big	0.29%	0.13%	-0.03%	-0.30%	0.73%	Big	1.233	0.463	-0.088	-1.007	2.866

Table 16 shows the alpha values of the five-factor model portfolios sorted on size-momentum, size-profitability, and size-investment. As the four-factor model, adding profitability and investment reduces the problems the three-factor model has with high momentum portfolio's slightly, but four out of five intercepts in these portfolios are still significantly different from zero (eight significant intercepts in total), and the problem with small portfolios are not reduced at all. The average alpha is not reduced either and is still relatively high with 0.35%. As earlier, it is difficult to find a pattern in the size-profitability, and size-investment sorted portfolios, where there are zero significant intercepts in the profitability sorting with an average alpha of 0.16% and one significant alpha in the investment sorting with an average alpha of 0.17%. This is similar to what was observed when adding momentum to the three-factor model, and the two models, therefore, perform very similar judging by the evidence presented so far. However, the four-factor model seems to describe the troublesome size-momentum sorting a little bit better.

Table 16 - Five-factor model regression for alternative sorts

Five Factor Model Regrssion

Sort on Size - Momentum (5x5)

	α						t(α)				
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	1.17%	0.58%	0.59%	-0.03%	-0.82%	**Small**	3.888	2.290	2.567	-0.132	-2.471
2	0.05%	0.52%	0.33%	-0.29%	-1.01%	**2**	0.198	2.367	1.658	-1.160	-3.841
3	0.14%	0.26%	0.17%	-0.23%	-0.69%	**3**	0.543	0.982	0.726	-0.909	-2.213
4	-0.02%	0.13%	0.15%	-0.14%	-0.72%	**4**	-0.074	0.451	0.602	-0.452	-2.302
Big	0.02%	0.18%	0.27%	0.18%	-0.02%	**Big**	0.067	0.474	0.918	0.656	-0.060

Sort on Size - OP

	α						t(α)				
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	0.31%	0.24%	0.27%	0.19%	0.38%	**Small**	1.345	0.961	1.113	0.697	1.098
2	-0.08%	0.09%	-0.04%	0.02%	-0.28%	**2**	-0.396	0.413	-0.187	0.099	-1.135
3	-0.33%	-0.04%	-0.04%	0.01%	-0.11%	**3**	-1.329	-0.153	-0.159	0.036	-0.402
4	-0.04%	0.03%	-0.16%	-0.18%	-0.23%	**4**	-0.154	0.123	-0.556	-0.735	-0.801
Big	0.13%	0.27%	-0.07%	0.33%	0.22%	**Big**	0.345	1.119	-0.283	1.364	1.195

Sort on Size - Inv

	α						t(α)				
	Low	**2**	**3**	**4**	**High**		**Low**	**2**	**3**	**4**	**High**
Small	0.21%	0.25%	0.42%	0.04%	0.47%	**Small**	0.971	1.062	1.605	0.163	1.318
2	-0.02%	-0.01%	0.03%	-0.08%	0.00%	**2**	-0.084	-0.043	0.131	-0.332	0.021
3	-0.27%	-0.17%	0.04%	-0.05%	-0.31%	**3**	-1.208	-0.699	0.174	-0.212	-1.222
4	0.14%	-0.15%	-0.14%	-0.16%	-0.15%	**4**	0.468	-0.503	-0.535	-0.594	-0.571
Big	0.17%	0.18%	0.55%	0.08%	0.08%	**Big**	0.497	0.538	2.441	0.376	0.333

6.6 *GRS* Test

Table 17 displays the results of the *GRS* test for the observation period 2006 - 2017. The *GRS* test was introduced by Gibbons, Ross and Shanken (1989) to investigate the overall fit of models. It tests whether the expected values of all intercepts of a group of portfolios are jointly zero. In this specific case, that means whether the expected values of all 25 portfolio intercepts are zero. For reasons of readability, only the key factors of interest are included in Table 17. For the size-*B/M* sorting, even the *CAPM* regression seems to deliver a low average intercept and a low test statistic. Nevertheless, the tests on single-factor regressions reject the null hypothesis of all intercepts being zero quite easily (98.8 %-level on average). The multi-factor regressions show even lower *GRS* test statistics as well as lower average absolute alphas (all below 1% on a monthly base), and the model that seems to explain the most is the five-factor model with the lowest *GRS* value (1.613), the highest p-value (5.03%) and one of the lowest average alpha values (0.22%).

Table 17 - GRS statistics on different portfolio sort based regressions

	GRS	**p-value**	**A|α|**
Size-B/M Portfolios			
Market	1.9057057	0.012643	0.00711
SMB Market	1.7764738	0.023535	0.002846
SMB HML	1.7919187	0.021893	0.00936
Market SMB HML	1.7827341	0.023003	0.002453
4-Factor	1.6221767	0.048041	0.002374
5-Factor	1.6127252	0.050322	0.002216
Size-Momentum Portfolios			
Market SMB HML	2.7624784	0.00017	0.003489
4-Factor	2.5428018	0.000537	0.003363
5-Factor	2.5005567	0.000679	0.003487
Size-OP Portfolios			
Market SMB HML	1.1800374	0.275508	0.00212
4-Factor	1.0558737	0.406415	0.001524
5-Factor	1.1540255	0.300784	0.001647
Size-Inv Portfolios			
Market SMB HML	1.3996246	0.122521	0.002379
4-Factor	1.3328047	0.159377	0.001948
5-Factor	1.1206629	0.33459	0.001669

The *GRS*-test rejects the null hypothesis of all intercepts being zero for the size-momentum sorting for all the three models mainly tested in the thesis on a high significance level (>99.1% level). This is expected based on the problems observed in previous sections for the different models, particularly in the small cap and high momentum portfolios. For the size-profitability and size-investment sorts, all the models fail to reject the null hypothesis, and especially the four- and five-factor model exceeds the 5% threshold by large margins. Overall, the four- and five-factor models perform similarly and a little bit better than the three-factor model. They all cover the observed returns in a convincing way when the portfolios are sorted on size and respectively *B/M*, profitability and investment, but all the models struggle with the size-momentum sorting.

The final five-factor regression and the *GRS*-tests conclude the analysis part of this thesis. Overall, the findings in the Chinese A-share market between 2006 and 2017 indicate a strong and significant size factor, a strong but insignificant market factor, a weak, negative and insignificant *B/M* factor, a significant negative medium-term momentum factor (reversal-effect), and a lack of profitability and investment effects. As for the overall factor models, the market and the size factors hold much explanatory power, whereas the explanatory power held by *B/M*, profitability and investment is questionable. The momentum effect seems to add some explanatory power, and it does not seem to be much that separates the performance of the four- and five-factor model. However, they both seem to explain the observed returns to a higher degree than the three-factor model. It is also the size-momentum sorting that causes the biggest problems for the models investigated in this thesis, particularly in the small cap and high momentum portfolios where all significant alphas are observed. These factor effects will be further discussed in section 8, but first the thesis moves on to testing the robustness of the effects observed in the analysis in the next section.

Section 7 – Robustness

Given the wide variety of results and methods applied in previous literature on factor effects in the Chinese market, an important part of this thesis is a thorough robustness check of the obtained results from the empirical analysis. Several checks are therefore performed. The robustness checks can be seen as an extension of the analysis, and will together with the analysis serve as the foundation for the discussion part in the next section. First, the period of investigation is changed to see how sensitive the previous findings are to alterations in the period of investigation. Second, the effect of using alternative measures of the market factor, the size factor, the book-to-market (*B/M*) factor, the momentum factor, and the risk-free rate is investigated. Third and last, value-weighted returns used throughout the paper are switched with equally weighted returns.

7.1 Period of investigation

Hu et al. (2018) suggest that the differing results in factor research on the Chinese market stem from the use of different sample periods. Furthermore, several changes to market regulations since the opening of the stock market in the early 1990's make it highly relevant to investigate the effect of changing the period. In particular, this section investigates the robustness of the significant variables from the analysis, in addition to the market and value factors. The economic returns to the two latter factors are high even though they are not significant. To assess how robust the findings in this paper are to different periods of investigation, the period is both prolonged as well as changed around within what the short history of the Chinese stock market allows. Table 18 displays factor premia for different periods of investigation.

Extending the period to cover returns from July 1996 to June 2017 provides some changes to the risk premia. The magnitude of both the market and the size premium decrease slightly, but both have delivered high positive returns over the period. The size premium is no longer significant on the 5%-level, but it is still significant on the 10%-level. However, the effect on the value and the momentum premia is more interesting. Both are heavily reduced in magnitude, and the negative value premium is very close to zero with a -0.15% monthly average return (-1.8% annually). The overall picture for all the periods tested seems to be that the market and size premia are relatively stable, while value and momentum are more sensitive to changes in the period under investigation. The market premium is large, positive and insignificant for all periods, whereas the size premium is large and positive for all periods and significant for three out of five periods that cover some part of the main period of investigation (2006-2017).

The negative value premium changes more, with one of the estimations being positive and the magnitude of three of them being considerably lower than the estimation for the period from 2006 to 2017. At the same time, the only *HML* premium that is positive is the one from 1996 to 2005. That is outside the main period investigated in the thesis. This period is excluded from the period of investigation in the analysis due to the belief that the period before the share-structure reform in 2005 is too different from the period after 2005, as elaborated on in section 5. The rest of the *HML* premia are negative and insignificantly different from zero, so at least the lack of a value premium seems to be robust.

The negative return on the momentum factor (*WML*) is even less robust to changes in the period of investigation than the *HML* factor. *WML* also has a small positive return in the period from 1996 to

2005, while the returns from the other periods are negative. However, the returns to the *WML* factor do vary a lot more depending on the period investigated. In fact, the highest absolute return (-0.71% monthly) is in the main period investigated in this thesis, while the other return values for the strategy are much closer to zero. Just moving the start of the period to 2009 instead of 2006 for example, the absolute value is a much lower -0.22% per month. That is far from being significantly different from zero, and that is the case for four of the six periods investigated here. There is only one other period with negative returns significantly different from zero in addition to the main results. The negative momentum effect over the medium-term, or in other words a medium-term reversal effect, does therefore not seem robust to changes in the time. As for value, the lack of a momentum effect is on the other hand evident.

In addition to the individual factor effects, the three-factor model is regressed on the size-*B/M* sorting to see if there are major changes. There is one more significant alpha in the period from 1996 to 2017, and the R-squared is slightly lower. However, it is expected that the explanatory power of the model falls a bit when the period reaches further back since more shares were non-floating in the earlier days of the market. At last, the coefficients on the risk premia do not change much either. It is also worth noting the very different return patterns and explanatory power in the period before 2005 compared to periods largely containing years after 2005. The period before 2005 is the only period with positive *HML* and *WML* premia, the market and size premia are much smaller than for the other periods investigated, and the explanatory power of the three-factor model is way lower than for the other periods. Gan et al. (2013) also find much lower R-squared values for this period. Their R-squared values range from 0.32 to 0.52. This supports the earlier argumentation in this thesis of the Chinese market before the many reforms around 2005 being too different from the last years' market to be part of the estimation of factor effects.

Table 18 - Factor risk premia using different observation periods

	RM-RF	SMB	HML	WML
2006-2017	0.97%	0.86%	-0.54%	-0.71%
t-value	1.25	2.13	-1.42	-2.24
1996-2005	0.37%	0.16%	0.37%	0.07%
t-value	0.51	0.39	0.77	0.14
1996-2017	0.81%	0.54%	-0.15%	-0.35%
t-value	1.57	1.84	-0.5	-1.2
2002-2017	0.61%	0.36%	-0.31%	-0.31%
t-value	1	1.05	-1.06	-1.08
2005-2017	1.14%	0.82%	-0.53%	-0.67%
t-value	1.58	2.04	-1.52	-2.06
2009-2017	0.51%	0.90%	-0.63%	-0.22%
t-value	0.67	2.46	-1.36	-0.59

7.2 Alternative definition of the explanatory variables

Since much of the comparable research uses different definitions of the variables used to form the dependent variable portfolios, it is necessary to assess whether the findings of the analysis are sensitive to alternative definitions of these variables. As in the previous section, only the market, size, value and momentum factors are investigated. Table 19 displays the risk premia for each of the four variables, as well as the R-squared values in the three-factor model. First, the market factor constructed in the analysis is changed with the return of the MSCI A Share Onshore Index. That makes the excess return on the market slightly higher (1.11%). However, the market factor is still not significantly different from zero due to its high volatility. The rest of the risk premia does not change, and the average R-squared value is slightly higher.

Secondly, size portfolio formation is done based on the total market value in the analysis. In this section, the total market value is changed with the floating market value. That is, market value as the basis for splitting the sample into size quintiles is calculated as price multiplied with the number of floating A-shares instead of total number of shares. Both approaches are common in the literature. All premia except the market premium increase slightly in absolute value when the floating market value is used. Size increases with 0.28% to 1.14% monthly average return, while *HML* and *WML* increase respectively 0.08% and 0.16% in absolute value. Furthermore, four alphas are significantly different from zero in the three-factor model, and the R-squared remains at an average of 0.92. The changes are all relatively small, and none of them changes the total picture, but it is worth noting that all of the premia increase in magnitude when floating market value is used to define the size of the firm. That does not only indicate that research findings of risk premia magnitude could be dependent on the definition of size, but also that the floating ratio has had some influence on returns over the period of investigation. The floating ratio as the share of non-floating shares of total shares still seems to have an impact, even though a big share of non-floating shares was floated more than a decade ago.

Third, the estimation of the size of a firm as the market value at the end of June every year might be wrong if there are substantial fluctuations in market value. If the time of observation is at a particularly bad time as a top or a bottom, the measured size then is not representative for the actual size of the firm. To check the robustness of the results to this, the size variable is defined as the average market value from July to June in the year before excess returns are calculated from July to June. That is, market value is estimated as the average market value from July of year t-1 to June of year t, and the excess returns are still estimated from July of year t to June of year t+1. The row named "Average M" in Table 19 the risk premia for this average constructed size variable. It shows that there are only minor changes in the risk premia and the overall interpretation does not change. The results are therefore robust to this alteration.

Fourth, the *B/M* ratio is constructed by dividing the book value of equity by the total number of shares and then the price per share in the analysis. In this section, the total number of shares is changed with the floating number of shares. That is, instead of dividing the book value of equity by total number of shares and then the price per share, the book value is divided by the number of floating shares and then the price per share. This is the equivalent of book value divided by floating market value. The results from this robustness test are quite interesting, as the returns from the *HML* factor turns slightly positive. It is not much larger than zero and not close to being significant on any meaningful statistical level, but it tells something about the robustness of the negative factor. Besides the changed *HML* factor, there are only minor changes to the other premia. Furthermore, the R-squared becomes slightly higher, as it increases from 0.92 to 0.93.

Fifth, the Carhart momentum factor used in the analysis is switched with the main momentum factor in Jegadeesh and Titman (1993). Following their approach, the momentum variable in month t is formed as the simple average over the months *t-1 to t-6*. The variable is kept constant for the next six months (t until t+5). This implies half yearly rebalancing of the variable instead of the monthly rebalancing done in the analysis, and a six-month formation period instead of 12 months. The row denoted “Momentum” in Table 19 lists the different factor premia, with their t-values in the row below. It can be observed that the absolute monthly returns on the momentum factor increase to a larger absolute value compared to its value in the analysis, as the negative monthly average returns increase from -0.71% (-8.90% annually) to -0.92% (-11.6% annually). This is large economically and means that a medium-term reversal strategy has earned an annual average of 11.6% over the period. However, due to the increased standard deviation of the new factor, it is no longer statistically significant on the 5%-level. Even though the significance level falls, the economic return is large, and it is still significantly different from zero on the 10%-level. At last, a momentum variable in the spirit of Carhart (1997) is constructed based on a shorter estimation period. Specifically, the estimation period is set to three months skipping the last. That yields momentum returns of -0.91%, which is highly significant with a t-value of -2.77. The medium-term reversal returns are therefore found to be robust to changes in variable construction.

The sixth and last robustness test for the construction of the variables that is performed is that the 3-month bank deposit rate is switched for the 3-month central bank bill. The risk-free rate is normally required to be traded in the market, but it is common in the Chinese factor model literature to use the bank deposit rate due to the short issuance history of central bank bills in China. The central bank bill could have been used for the entire main period examined in this thesis, but the bank deposit rate is used for complete comparability to earlier periods. The line called *RF* in Table 19 also shows that there is practically no difference between using the two different rates for the factor premia and the regressions. The only observable difference is that the market premium is slightly lower.

Table 19 - Risk premia for different variable constructions

	RM-RF	SMB	HML	WML	Avg. R^2
Analysis	0.97%	0.86%	-0.54%	-0.71%	0.92
t-value	1.25	2.13	-1.42	-2.24	
MSCI RM	1.11%	0.86%	-0.54%	-0.71%	0.93
t-value	1.35	2.13	-1.42	-2.24	
Float MV	0.97%	1.14%	-0.62%	-0.87%	0.92
t-value	1.25	2.82	-1.64	-2.7	
Mean MV	0.97%	0.71%	-0.59%	-0.91%	
t-value	1.25	2	0.64	-2.23	
B/M	0.97%	0.81%	0.22%	-0.71%	0.93
t-value	1.25	2	0.64	-2.23	
Momentum	0.97%	0.86%	-0.54%	-0.92%	
t-value	1.25	2.13	-1.42	-1.89	
RF	0.93%	0.86%	-0.54%	-0.71%	0.92
t-value	1.2	2.13	-1.42	-2.24	

7.3 Equally weighted returns

Next, the value-weighted returns used throughout the paper are changed to equally weighted returns. Equally weighted returns are useful to test if the main results remain solid. As before, only the market, size, value and momentum factors are investigated. Table 20 displays the comparison of the two methods, and the only premium that changes considerably is the excess market return. It more than doubles and becomes highly significant. Besides that, the size premium drops a bit but remains highly significant, the *B/M* premium rises a bit in absolute value but remains insignificant on the 5%-level, and average momentum return stays exactly the same as before. The overall picture does therefore not change much.

Table 20 - Comparison value weighted and equal weighted returns

	RM-RF	SMB	HML	WML
Value w	0.97%	0.59%	-0.54%	-0.71%
t-value	1.25	2.13	-1.42	-2.24
Equal w	2.26%	0.59%	-0.65%	-0.71%
t-value	2.41	2.44	-1.89	-2.32

That concludes the robustness checks performed on the main results. The overall picture is that the size and market factor seem to be robust to changes in the period of investigation and variable construction, while the *B/M* and momentum factors appear to be more fragile. None of the *HML* premia for different periods of investigation was significantly different from zero, and when the factor is constructed differently the sign of the factor changes to positive. However, the lack of a value premium does seem apparent. The momentum premium also becomes insignificant for most of the other periods investigated, and it, therefore, seems like the momentum pattern found in the main period of investigation might not be found generally. As for the value premium, the lack of a momentum premium over the medium-term seems evident. At last, the explanatory power of the different factors in the three-factor model does not change much in any of the robustness checks performed in this section.

Section 8 – Discussion

So far, this thesis has found strong and robust market and size effects, weak value, profitability and investment effects, and a negative momentum effect in the Chinese A-share market. In addition, the three-factor model explains more of the return variability in Chinese stock returns than the *CAPM*, and the four- and five-factor model are even better fits. This section will discuss all of these findings, and attempt to place them in the context that is built throughout the thesis with the Chinese market overview and the literature review as background.

Equity premium

The market premium is highly positive with a monthly excess return of 0.97% (12.2% annually) but is not significant. This is expected based on the Chinese factor model literature, as all papers reviewed in this thesis find positive excess returns for both the market factor. Most of the papers on the Chinese A-share market also find a high but insignificant market premium due to the high volatility of the market. Examples are Cheung et al. (2015), Chen et al. (2015) and Hu et al. (2018). Fama and French (1993) also report a positive but insignificant market premium for the U.S. market, something they attributed to a well-known problem with research on equity returns; the previously mentioned high volatility of said returns.

Size

The size premium is significant at the 5%-level and has yielded a 0.86% monthly average return (10.89% annually) over the period from July 2006 to June 2017. Returns also monotonously decrease in size for both the univariate and the multivariate sorting. Regardless of which of the other variables size is combined with, the factor loadings on the *SMB* factor in all regressions are large, and it contributes greatly to the explanatory power added to the *CAPM* regression. As stated above, all research reviewed in this thesis find a size premium in the Chinese A-share market, and mostly the premium is significant. Examples of this are Chen et al. (2015) and Hu et al. (2018).

One example of a paper that did not find a significant size premium is Cheung et al. (2015). They attribute this to the high volatility in their smallest portfolios, as well as differing methodology from other research on the Chinese market. Their methodology differs in the way that they only investigate the constituents of the MSCI China A Share Index, in contrast to most other research on the A-share market that investigates all shares listed on the Shanghai and Shenzhen Stock Exchange. MSCI has size and liquidity filters they screen the stocks they include in their index on (MSCI Global, 2018), which typically excludes the stocks that are categorized as small in the total A-share universe. Fama and French (1993) find an insignificant size premium in the U.S. market. However, they still find it to be important due to the high factor loadings, which is also the case in this thesis.

Relevant explanations for the size premium are offered by both Banz (1981) and Chan and Chen (1991). Banz (1981) suggests that there could be a lack of available information on small firms, which means that they have to pay an additional premium due to estimation risk. It is likely that small firms have to pay a premium in a market like China where studies have shown signs of both dubious reporting practices and high level of tunneling (Drew, Naughton, & Madhu, 2003). Especially smaller firms with less publicly available and less reliable information could be subject to this. That could also explain why Cheung et al. (2015) find no significant size premium, contrary to all other research examined in this thesis, as they only investigate the constituents of the MSCI China A Share Index. The MSCI constituents are included in the index because they live up to standards related to investment and liquidity, and for that reason, the information on these firms should be of good quality. That there is a significant size effect when the whole market is investigated could indicate that insufficient information and thereby estimation risk is a source of the strong size effect in China.

On the other hand, the size effect in China might be boosted by the particular special treatment system the regulatory authorities enforce. Distressed and badly performing firms are rarely delisted but instead receives what is referred to as "Special Treatment." Instead of delisting the distressed firm another firm seeking listing takes over their listed "shell." This means that badly performing firms might drag down the bigger size quantiles, but when they become so small that they would be categorized as a smaller firm they might disappear without dragging down the smaller size quantiles and be replaced by a more competitive firm. This firm is then more likely to drag the performance of small firms up than the old, badly performing firm. Thus, the returns of small firms might be kept artificially high by the special treatment system. The issue of survivorship bias is therefore highly relevant in the Chinese market.

An aspect that could drag the size premium in the other direction is the tough listing requirements in the Chinese market. The listing requirements have been shown to cause large post IPO drops in performance because firms exhaust their resources to achieve public listing[70]. Assuming the firms that enter the market are smaller than most current listings, this post IPO drop would drag the performance of small firms in the opposite direction of the previously discussed "Special Treatment" effect. Both of these regulations, as well as the possible estimation risk explanation and their effect on the small firm returns in the Chinese-A share market are interesting topics for future research.

Fama and French (1992) explain their initial findings with the observation of Chan and Chen (1991), which suggest that small firms do not earn a premium solely because they are small, but rather because the group of small firms is highly populated by marginal firms. That is, firms with low efficiency and high leverage. Due to that, they are also more exposed to shifting economic conditions. The summary statistics for the stock returns also suggest that small firms are less efficient than big firms. On the other hand, the leverage patterns look rather random. That is, there does not seem to be a strong relationship between size and leverage level except in the two quantiles with lowest *B/M* ratio. Therefore, there is no clear indication one way or the other for this explanation based on the findings in this thesis.

As a last note on the size factor, much of the research uses only the floating market value to define the size of a firm. That might be misleading in China since significant amounts of shares are not listed on any of the stock exchanges. These stocks are still a part of the firm's equity, and this thesis thus argues that they should be taken into consideration when defining the size of a firm. The robustness checks performed in section 7 also indicate that the premia might be higher in absolute value when the floating market value is used instead of the total market value.

The Chinese size factor has been found to be strong and robust in this thesis. Possible explanations such as estimation risk due to greater information uncertainty for smaller firms and regulatory features

[70] Allen et al. (2017)

such as the special treatment system causing a survivorship bias in the data have been presented. These explanations are interesting, but the empirical evidence in this thesis is insufficient to conclude one way or another. However, the topic is highly relevant for future research. In the extension of this, it will also be interesting to see how the size effect develops in China going forward. That could shed some light on the disappearance of the size effect in global markets. In particular, whether or not it persists could indicate if its initial disappearance was caused by overexploitation of the initial pattern of abnormally high returns to small stocks.

Value and profitability

In contrast to the size factor, the value factor has earned -0.54% monthly average return (-6.26% annually) over the period from July 2006 to June 2017. That is, growth firms have outperformed value firms over the period of investigation. However, the outperformance of growth firms is not close to being significantly different from zero, and the excess return pattern for both the univariate and the multivariate sorts are more mixed than for the size factor. In addition, the regression factor loading on the *HML* factor in all regressions are rather small compared to the market and size. It is also found in section 7 that the *HML* factor is not robust towards changes in variable construction or period of investigation.

Nevertheless, the mean over the period is negative, which is unusual especially compared to international markets, but also in more recent research on the Chinese market. The *B/M* ratio has maybe been the strongest and most reliable explanatory factor for stock returns for decades and across several markets, with evidence of it provided in widely recognized papers as Fama and French (1992, 1993, 1998, 2012) and Asness, Moskowitz and Pedersen (2013). However, research on the later Chinese market is more conflicted. For example Cheung et al. (2015) report a significant positive value premium in the period from 2002-2013, Hu et al. (2018) report an insignificant positive value premium in the period from 1997 to respectively 2013 and 2016, and Xu and Zhang (2014) and Belimam (2018) actually report a slightly negative value premium over periods starting in 2007 or later. In addition, research on the Chinese market before the floating reform in 2005 reports a negative value premium. Both Drew et al. (2003) and Gan et al. (2013) report negative value premia.

The profitability factor has an average return of -0.28% per month (-3.35% annually). However, even though the average return to the factor is negative, the economic magnitude of the return is small, and it is far from being significantly different from zero. There is therefore rather a lack of a profitability premium than a negative premium. Following Miller and Modigliani (1961), Fama and French (2015) argue that higher expected future earnings imply a higher expected return. As seen, this is not the case in the Chinese market. Profitability factor returns very close to zero are also reported in comparable research, with Lin (2017) reporting average monthly returns varying from -0.009% to 0.021% between 1997 and 2015 and Belimam (2018) reporting an average monthly return of 0.18% from 2011 to 2016.

Both Drew et al. (2003) and Wang and Xu (2004) offer several explanations for the negative value premium they observe in the early Chinese market. First, overexploitation of the value effect is presented as an explanation. That is, many trade to exploit this well-known pattern, thereby driving the price up and reducing future returns. A second explanation offered is irrational investor behavior. The *B/M* ratio proxies for fundamental value, and if the book value of a firm is difficult to assess or investors do not pay much attention to fundamentals the *B/M* ratio does not carry much informational value for pricing. This explanation is equally valid for the profitability factor since it is also a factor related to fundamentals.

Research on the early market reported sketchy accounting practices, but it is difficult to tell if this is still the case. Nevertheless, if the market has the impression that accounting data is not trustworthy or difficult to assess, the trust in fundamentals could be gone, and it might lose its value regardless of whether or not the impression is true. The explanation of investors not paying much attention to fundamentals is also interesting due to the large share of retail investors in the market. 80-90% of the market are reported to be retail investors, which means that most of the market consists of unsophisticated investors.

This explanation is very much linked to the perception of the Chinese market as a casino. However, this perception is contradicted by Carpenter et al. (2018). They argue that this is a false impression and that the Chinese stock market is highly linked to fundamentals. They argue that Chinese investors pay up for size, growth, liquidity and long shots, but similar to this thesis they also find that the *B/M* ratio is not a significant pricing factor. However, this is an area of disagreement, with for example Allen et al. (2017) claiming that the Chinese stock market is disconnected from its overall economy.

A third explanation offered is that special features of the Chinese market affect the returns. The big share of retail investors has been mentioned, but the share structure with floating and non-floating shares and government interventions could also be possible reasons. Early research on the market, in particular Wang and Xu (2004) successfully replace the *B/M* ratio in the three-factor model with the floating ratio. They find that the floating ratio worked well as a proxy for corporate governance. Even though the influence of the floating ratio declined after the share-structure reform in 2005, it could still be something that influences stock returns. Many particular features are still present in the market, such as the T+1 rule, frequent trading halts, and frequent stock suspensions, and even though short selling has been legal since 2006, it is reported to be difficult. The effect of all of these elements could be a less efficient market that incorporates less fundamental information.

Similar to the size factor, the *B/M* factor can be constructed in two different ways. It can be constructed the traditional way as the book value of equity divided by floating market value, or it can be constructed as the book value of equity divided by total market value. When the *B/M* definition is changed in the robustness checks in section 7, the sign of the *B/M* ratio changes. That is, it turns from negative to positive. First of all, that indicates that the found negative *B/M* premium is not robust, as other robustness tests also show. Second, it indicates that sign and magnitude of the *B/M* premium must be interpreted in the light of how the risk factor is constructed.

The value and profitability factors have both yielded negative average returns over the period investigated. However, none of them are significantly different from zero, and the value premium is neither robust towards change in variable construction nor period of investigation. This is consistent with the most recent factor literature on the Chinese market. One reason for the lack of value and profitability premia is suggested to be lack of attention to fundamentals due to unreliable accounting figures or the high amount of retail investors in the market. Another explanation put forth is the high amount of market peculiarities in the Chinese market. Nevertheless, the lack of *B/M* and profitability premia are interesting held up towards other markets, and are interesting areas for future research.

Investment

The investment factor has earned an average monthly return of 0.19% (2.28% annually) in the Chinese A-share market from 2006 to 2017, which is low in economic terms and far from being statistically significant. Fama and French (2015) rationalize the investment factor by showing to the Miller and Modigliani (1961) equation which proposes that higher expected growth should imply lower expected return. An explanation to this premium is put forth by Titman et al. (2004), which state that investors dislike the "Empire building" attitude some managers have and that they punish the firm by pushing its stock price down. The existence of such a premium could make sense in the Chinese market, which Allen et al. (2017) find to have low investment efficiency compared to other markets as well as foreign-listed Chinese companies. Despite this, the investment factor is shown to be weak.

However, the low investment returns are consistent with findings in recent research on the Chinese market. Lin (2017) reports investment premia between 0.03% and 0.044% in the Chinese market between 1997 and 2015, while Belimam et al. (2018) report an investment premium of 0.05% between 2011 and 2016. Lin (2017) also finds the investment factor to be redundant in the five-factor model and offers two possible explanations for the lack of investment effects in China. First, he argues that the bank-oriented Chinese market could make profitability a better predictor of future performance than investment. Second, he shows to La Porta et al. (1999, 2002) and Yu et al. (2010), arguing that firms in developing countries are more likely to have an ownership that pressure management into pursuing the owner's personal benefits. This weakens the predictive power of the investment factor.

Based on observed patterns and the estimated premium in this thesis, as well as comparable research, the investment factor seems to be weak in the Chinese A-share market.

Momentum

A momentum strategy based on recent winners minus recent losers over the last 12 months minus the last has earned -0.71% monthly average return (-8.23% annually) over the period from July 2006 to June 2017. That is, contrary to the conventional momentum strategy over the medium term, a medium-term reversal strategy has been profitable over the period of investigation. This reversal strategy is also significantly different from zero on the 5%-level and is also robust towards change in variable construction. However, the return patterns for the univariately and multivariately sorted portfolios are not clear and rather random. Furthermore, the robustness checks in section 7 show that the variable is not robust towards changes in the period of investigation. The total picture for the reversal returns are therefore mixed, but a reversal strategy has yielded significant returns over the period.

Cheung et al. (2015) state that they are puzzled over the failure of the existing literature to detect a momentum premium in China, with the apparent short-term nature and high turnover of the A-share market. However, they are not able to detect such a pattern either. The failure to detect a momentum factor in Chinese stock returns is therefore not groundbreaking, but the statistically significant negative premium stands out. However, Cheung et al. (2015) suggest that the momentum strategy seems to be highly sensitive to the time horizon, which is an appealing explanation based on the evidence presented in this thesis. First, the momentum factor is constructed as a simple average that moves over time, so timing it wrong in volatile markets might lead to doing the exact opposite of what would earn profits. That is, the signal would turn positive after a period of positive returns, but around that time the period of momentum actually ends and the stock reverts down again. At the same time, the robustness checks in section 7 show that reducing the estimation period to three or six months does not change anything.

Second, the observed momentum pattern is sensitive to changes in the period of investigation. The time period robustness checks in section 7 show that the negative momentum returns are reduced to only half of the main result in the years between July 1996 and June 2017. Over this period the returns from a strategy of 10% biggest winners minus 10% biggest losers actually turn to slightly positive (Appendix). In fact, the negative momentum returns are significantly reduced for four of the five other time periods investigated, and the main findings are the highest absolute value of all. In addition to the main findings, there is only one other momentum premium that is significantly different from zero.

It can also be seen in the Appendix that the market factor in the main period under study is more volatile than all of the other periods. This is interesting because momentum has been shown to be negatively related to volatility (Koesterich, 2017). Table 21 shows the return to the momentum strategy for every period investigated in the robustness section, and momentum returns do indeed seem to be decreasing in market volatility. Except from the slightly lower volatility from 2009 to 2017 than from 1996 to 2005, higher market volatility is constantly associated with lower returns to the constructed momentum factor. This is an interesting observation for what might cause momentum returns, or the lack of momentum returns, not only in China but also in other markets. It indicates that further research on the relationship between volatility and momentum returns might be fruitful. Furthermore, a volatility factor might carry a risk premium in the Chinese A-share market. Evidence of the latter is already reported in Cheung et al. (2015) and Chen et al. (2010).

Table 21 - Return of momentum factor for different observation periods

	Mom r	Std RM-RF
96 to 05	0.07%	7.57%
09 to 17	-0.22%	7.45%
02 to 17	-0.31%	8.17%
96 to 17	-0.35%	8.22%
05 to 17	-0.67%	8.68%
06 to 17	-0.71%	8.90%

From the results presented in this thesis, it is evident that there has been no momentum premium over the standard medium-term. The analysis shows a significant reversal premium over the medium-term, and that is robust to changes in variable construction. However, the premium is not robust towards changes in the period of investigation. Another interesting observation is the link between market volatility and returns to the momentum factor. Volatility has previously been shown to deliver significant returns (Cheung (2015)), and both the volatility factor in the Chinese market and the link between momentum returns and market volatility should be interesting topics for future research.

Factor models

Even though many of the factor effects discussed so far have not delivered significant returns in the Chinese market as they have in other markets, the factors could still be an important part of an overall description of the market. Fama and French (2015) state that multivariate regressions measure the marginal effect of each factor holding the other factors constant. This marginal effect does not only depend on the individual characteristics of a factor, but also on the correlation it has with other factors. The overall fit of the model is therefore also dependent on the interaction with the other factors in the regressions.

All of the multifactor models are better at explaining equity returns in the Chinese market than the *CAPM*. This can be seen from average R-squared and alpha values, as well as the *GRS* test in the analysis. The average R-squared for the *CAPM* regressed on the size-*B/M* sorted dependent variable portfolio returns is 0.74, while the average R-squared for the three multi-factor models investigated in this thesis is approximately 0.92 across the four different sorts performed. Also, the three multi-factor models fail to reject the null hypothesis of all intercepts jointly being zero in the *GRS* test for three of four sorts, while the *CAPM* clearly rejects the null hypothesis. This is in line with all papers reviewed in this thesis.

Furthermore, judging by the *GRS* scores of the three-factor model and a model only including the market and size factors, the model only including market and size is a better fit in the Chinese market. Recall from section XX that the three-factor model had one significant alpha in the large-cap, high *B/M* portfolio. Table XX shows the alpha values of the size-*B/M* sorted portfolio returns regressed on the model with the market and the size factor. It is worth to notice that the significant alpha in the three-factor model is no longer significant. None of the alphas in this model are significantly different from zero for the size-*B/M* sort. On the other hand, the absolute average alpha is quite a lot higher for the model only including market and size, so the evidence on the relative strength of these two models is mixed. This conflicts with the initial findings of Fama and French (1993) on the U.S. market, where the *B/M* factor plays an important role. However, in some way it is in line with the literature on three-factor models in the Chinese A-share market since the impact and importance of the *B/M* ratio in the overall models are the disagreement regarding the three-factor model. An example of a paper that does not find a significant *B/M* factor is Hu et al. (2018), while an example of the opposite is Cheung et al. (2015). Belimam et al. (2018) find the value factor to be redundant in the three-factor model they test.

Table 22 - Single factor regression on Sort - Momentum sort

Rp-Rf = α + [RM-RF] + SMB + e											
Sort on Size - Momentum (5x5)											
α						t(α)					
	Low	2	3	4	High		Low	2	3	4	High
Small	0,52 %	0,25 %	0,32 %	0,01 %	0,16 %	**Small**	1,345	0,871	1,294	0,035	0,629
2	0,10 %	0,20 %	-0,17 %	-0,36 %	-0,25 %	**2**	0,297	0,643	-0,733	-1,458	-0,929
3	0,10 %	0,07 %	-0,31 %	-0,48 %	-0,53 %	**3**	0,305	0,240	-1,192	-1,944	-1,755
4	0,37 %	-0,41 %	-0,41 %	-0,38 %	-0,29 %	**4**	1,007	-1,322	-1,274	-1,315	-0,882
Big	0,42 %	0,02 %	-0,20 %	-0,42 %	0,35 %	**Big**	1,207	0,072	-0,755	-1,208	0,921

Regardless of which of the two models discussed in the previous paragraph is the best model for the Chinese equity market, the four-factor model of Carhart (1997) outperforms the three-factor model on the main metrics the models are evaluated on in this thesis: average alphas and *GRS*-test scores are lower, and accompanying p-values are higher for all the four sorts performed in the analysis. This is in line with Fama and French (2015), which also find that the five-factor model outperforms the three-factor model on the main metrics the models are evaluated on. It is also in line with the findings of Lin (2017) for the period from 1997 to 2015, while Belimam et al. (2018) find the three-factor model superior from 2011 to 2016. The relative strength between the four- and five-factor model is harder to conclude on, as they lie very close to each other in *GRS*-score and average alpha for all the four dependent variable sorts. In fact, the difference in average alpha is not larger than 0.02% for any of the four sorts, and they both perform better than the other on average alpha in two out of the four sorts. However, the five-factor model performs better on the *GRS*-test in three out of four sorts. In that way, it could be claimed that the five-factor model is the superior model, but it should be noted that the outperformance is marginal.

The size-momentum sort is the only sort that causes problems for the models tested here, where the problem portfolios seem to be small-cap and high momentum portfolios. The addition of profitability and investment seem to mitigate these problems to the same extent as the momentum, as the *GRS*-test is slightly better for the five-factor model than the four-factor model for this sort. However, none of the tested models mitigate the problems to a great extent as there still is a considerable amount of significant alphas for both models and the null hypothesis of all intercepts jointly being zero is rejected on a very high confidence level for this sorting. Besides this, the only two portfolios with significant alphas are the large cap-high *B/M* and the large cap-third highest investment portfolio.

As a last note, it should also be mentioned that the different results in this thesis could be due to different input data. Most research done on the Chinese market uses data gathered from a Chinese database, typically China Stock Market and Accounting Research (CSMAR) or the Chinese Capital Market Database. This thesis gathers all the necessary data from *Datastream*, and even though Allen et al. (2017) state that a cross-check of *Datastream* and WIND/CSMAR stock and accounting data for Chinese A shares yield similar results, one cannot be entirely sure. They also state that they perform a data cleaning process based on these databases, which this thesis unfortunately cannot do due to lack of access.

Based on the accessible data and assuming that no or only minor computational mistakes have been made, the observed market and size effects are strong and in line with existing research. The value, profitability, and investment factors have not delivered significant returns over the period, and the profitability and investment returns are very close to zero. The value returns are negative and larger in absolute value than for profitability and investment, but they are not robust to changes in model construction and estimation period. The momentum factor is negative and significantly different from zero, and it is robust to changes in the estimation and holding period. However, it is not robust to changes in period of investigation. Furthermore, the four- and five-factor model outperform the three-factor model in the Chinese A-share market from 2006 to 2017, while the relative difference between the two is almost non-existing judging by the evidence presented in the analysis. Lin (2017) finds the investment factor to be redundant, and Belimam et al. (2018) find the value variable to be redundant. An interesting area for future research is the relative importance of especially the value, investment and profitability factor.

Section 9 – Conclusion

To answer how well factor models explain equity returns in the Chinese market this thesis runs a variety of models and robustness checks. After comparing standard measures of fit and the more comprehensive *GRS* tests with each other the thesis finds the four- and five-factor model to be the best fit. The individual effect of some of the factors might be less clear. However, the factors and their overall interplay seem to explain much of the variation in the Chinese equity market. For the multifactor models tested, three out of four sorts performed are unable to reject the null hypothesis of the *GRS* test, and the average R-squares are approximately 0.92.

Individual factor effects have also been a focus of the thesis. From the results presented in the analysis, it can be seen that the market premium is positive but insignificant, the *Size* premium is strong and significant whereas there seems to be a lack of a *B/M*, profitability and investment premia. Furthermore, a significant medium-term reversal effect is present in the market rather than a momentum effect. These findings contrast comparable research on global markets but are broadly in line with research on the Chinese market.

To verify the robustness of the unveiled factor effects in the Chinese market, this thesis performs comprehensive testing. Neither changing the construction of the variables nor changing the time frame lead to major changes. Only the changed time frame cause some variation in the estimated premia. The findings in the previous literature, as well as the observations made in this thesis, also indicate that documented factor effects in different academic papers are highly dependent on the period of investigation.

As stated in the outline section, the first five sub-questions derived to answer the problem statement are rather directed towards building the foundation for the analysis. Nevertheless, the review of the Chinese market and the literature review lead to some interesting insights related to methodology. First, the existence of floating and non-floating shares in the A-share market creates a problem for the construction of both size and *B/M* variables. Second, the short history and many regulatory changes in the market cause a shorter time frame for estimation of the time series than research performed in other markets. Third, the availability and usage of data sources differs. There exist several databases that provide data for the Chinese equity market as well as Chinese accounting figures.

Overall, sources, methodology, and alterations to the standard (Fama and French) methodology in global markets are inconsistent across Chinese factor model research. Especially deviations stemming from the varying variable construction and differences between global data providers like *Datastream* and local Chinese databases harm the comparability and transparency. This makes it particularly hard to compare research and findings directly with each other. Based on what has been investigated on the Chinese market there is a need for an established common methodology at least for the construction of the variables/ factors. Therefore, the reader is advised to interpret the results in this thesis with caution.

Furthermore, the findings in this thesis are connected to several prevailing explanations of different risk premia in the discussion section, and many of these might be fruitful areas for future research. First, the reasons for the size premium are highly interesting due to what it might imply for size on a global scale. The strong Chinese premium is of course also interesting, but building an understanding of the reasons for it might reveal fundamental insights about the small size premium. Specifically, one would start by investigating whether market structures cause or enforce this factor. Second, the significant medium-term reversal effect contrasts findings in many other markets. Due to this, a detailed investigation into this specific factor is desirable. Third, based on Chinese market peculiarities and reviewed literature, dividend yield and volatility are factors that look promising but have not been investigated in this thesis.

For the overall factor models, the interplay between the factors investigated is a natural continuation of the investigation in this thesis. In addition, other factors that are important for stock prices in China might be present and should be investigated. Due to the explanatory nature of this thesis, the focus has rather been on building a solid framework for the statistical tests performed and the conclusions drawn from the findings. Therefore, the statistical tests can be expanded to further test the robustness of the observations. Two specific examples of that are the 2x2 and 2x2x2x2 independent variable sorts performed by Fama and French (2015), as well as the addition of further asset pricing evaluation metrics.

At last, research that contributes to an established methodology for factor research on the Chinese market is necessary. Especially on the topics of period of investigation, construction of variables and usage of different data sources this would be highly beneficial for reasons of comparability across literature.

Bibliography

Aharoni, G., Grundy, B., & Zeng, Q. (2013). Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. *Journal of Financial Economics, 110(2)*, 347-357.

Allen, F., Qian, J., Shan, C., & Zhu, J. (2017). Dissecting the Long-term Performance of the Chinese Stock Market. 1-70.

Ang, A. (2014). *Asset management: A systematic approach to factor investing.* Oxford University Press.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance, 68(3)*, 929-985.

Bai, J., Fleming, M. J., & & Horan, C. (2013). *The microstructure of China's government bond market.* New York: Federal Reserve Bank of New York Staff Reports, no. 622.

Baker, N. L., & Haugen, R. A. (2008). *Case Closed.* The Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques.

Baker, N. L., & Haugen, R. A. (2012). Low risk stocks outperform within all observable markets of the world.

Ball, R. (1978). Anomalies in relationships between securities' yields and yield-surrogates. *Journal of financial Economics, 6(2-3)*, 103-126.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial Economics, 9(1)*, 3-18.

Barberis, N., Shleifer, A., & & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49(3)*, 307-343.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of financial Economics, 12(1)*, 129-156.

Belimam, D. T. (2018). An Empirical Comparison of Asset-Pricing Models in the Shanghai A-Share Exchange Market. *Asia-Pacific Financial Markets, 25(3)*, 249-265.

Bender , J., Briand , R., Melas, D., & Subramanian, R. A. (December 2013). *MSCI.com.* Hentet fra Foundations of Factor Investing: https://www.msci.com/documents/1296102/1336482/Foundations_of_Factor_Investing.pdf/004e02ad-6f98-4730-90e0-ea14515ff3dc

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The journal of Finance, 43(2)*, 507-528.

Black, F., Jensen, M. C., & Scholes, M. S. (1972). *The capital asset pricing model: Some empirical tests.* Praeger Publishers Inc.

Campbell, C. J., Cowan, A. R., & Salotti, V. (2010). Multi-country event-study methods. *Journal of Banking & Finance, 34(12)*, 3078-3090.

Campbell, J. Y., Lo, A. W., & MacKinlay, C. A. (2012). *The Econometrics of Financial Markets.* Princeton University Press.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 57-82.

Carpenter, J. N., Lu, F., & Whitelaw, R. F. (2018). The Real Value of China's Stock Market. 1-39.

Chan, K. C., & Chen, N. F. (1991). Structural and return characteristics of small and large firms. *The Journal of Finance, 46(4)*, 1467-1484.

Chan, L. K., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan. *Journal of Finance, 46(5)*, 1739-1764.

Chen, J. (2017). What explains the investment anomaly in the Chinese stock market? *Nankai Business Review International, 8(4)*, 495-520.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of business*, 383-403.

Cheung, C., Hoguet, G., & Ng, S. (2015). Value, size, momentum, dividend yield, and volatility in China's A-Share market. *Journal of Portfolio Management, 41(5)*, 57-70.

China Insights. (u.d.). Hentet 1. September 2018 fra UBS: https://www.ubs.com/microsites/china-insights/en/insights/impact.html

Claessens, S., Dasgupta, S., & Glen, J. (1995). *The cross-section of stock returns: Evidence from the emerging markets.* Washington: World Bank Publications.

Conrad, J., & Kaul, G. (1998). An anatomy of trading strategies. *The Review of Financial Studies, 11(3)*, 489-519.

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance, 52(1)*, 1-33.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance, 61(4)*, 1605-1643.

Daniel, K., Hirshleifer, D., & & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance, 53(6)*, 1839-1885.

Daniel, K., Titman, S., & Wei, K. J. (2001). Explaining the cross-section of stock returns in Japan: Factors or characteristics? *The Journal of Finance, 56(2)*, 743-766.

Datastream. (5. May 2018). Hentet fra http://share.thomsonreuters.com/assets/newsletters/ssr/Datastream.pdf

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, covariances, and average returns: 1929 to 1997. . *The Journal of Finance, 55(1)*, 389-406.

De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of finance, 40(3)*, 793-805.

Drew, M. E., Naughton, T., & Madhu, V. (2003). Firm Size, Book-to-Market Equity and Security Returns: Evidence from the Shanghai Stock Exchange. *Australian Journal of Management, 28(2)*, 119-139.

Eun, C. S., & & Huang, W. (2007). Asset pricing in China's domestic stock markets: Is there a logic? *Pacific-Basin Finance Journal, 15(5)*, 452-480.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33(1),*, 3-56.

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance 50(1)*, 131-155.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1), 55-84.

Fama, E. F., & French, K. R. (1998). Value versus Growth: The International Evidence. *The Journal of Finance*, 53(6), 1975-1999.

Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics, 82(3)*, 491-518.

Fama, E. F., & French, K. R. (2008). Dissecting Anomalies. *The Journal of Finance*, 63(4), 1653-1678.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics, 105(3)*, 457-472.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116(1)*, 1-22.

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of financial Economics*, 123(3), 441-463.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy, 81(3)*, 607-636.

Fama, E. F., French, K. R., & Davis, J. L. (2000). Characteristics, covariances, and average returns: 1929 to 1997. *The Journal of Finance, 55(1)*, 389-406.

Financial Times. (05. May 2018). *Financial Times*. Hentet fra End of the road for China's 'B' market: https://www.ft.com/content/254b3b6e-5a2a-11e2-a02e-00144feab49a

Financial Times. (u.d.). *End of the road for China's 'B' market*. Hentet 7. May 2018 fra Financial Times: https://www.ft.com/content/254b3b6e-5a2a-11e2-a02e-00144feab49a

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics, 111(1)*, 1-25.

Gan, C., HU, B., Liu, Y., & Li, Z. (2013). An empirical cross-section analysis of stock returns on the Chinese A-share stock market. *Investment Management and Financial Innovations, 1*, 127-136.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica 57*, 1121-1152.

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *The quarterly journal of economics, 118(1)*, 107-156.

Graham, B., & Dodd, D. (1934). *Security Analysis: Principles and Technique.* New York: Whittlesey house, McGraw-Hill book Company, Incorporated.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *The Review of Financial Studies, 15(3)*, 783-803.

Griffin, J. M., Ji, X., & Martin, J. S. (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *The Journal of Finance, 58(6)*, 2515-2547.

Hanauer , M. X., & Linhart, M. (2015). Size, value, and momentum in emerging market stock returns: integrated or segmented pricing? *Asia-Pacific Journal of Financial Studies, 44(2)*, 175-214.

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *The review of financial studies, 8(3)*, 773-816.

He, X., Li, M., Shi, J., & Twite, G. J. (2009). Determinants of dividend policy in Chinese firms: Cash versus stock dividends.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance, 54(6)*, 2143-2184.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies, 28(3)*, 650-705.

Hu, G. X., Chen, C., Shao, Y., & Wang, J. (2014). Fama-French in China: Size and Value Factors in Chinese Stock Returns. 1-41.

Hu, G. X., Chen, C., Shao, Y., & Wang, J. (2018). Fama–French in China: size and value factors in Chinese stock returns. *International Review of Finance*, 1-42.

Hu, G. X., Pan, J., & Wang, J. (2018). Chinese Capital Market: An Empirical Overview. *National Bureau of Economic Research*, 1-72.

Ince, O. S., & Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research, 29(4)*, 463-479.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48(1)*, 65-91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance, 56(2)*, 699-720.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American economic review 76(2)*, 323-329.

Johnson, T. C. (2002). Rational momentum effects. *The Journal of Finance, 57(2)*, 585-608.

Kang, J., Liu, M. H., & Ni, S. X. (2002). Contrarian and momentum strategies in the China stock market: 1993–2000. *Pacific-Basin Finance Journal, 10(3)*, 243-265.

Koesterich, R. (30. March 2017). *The seesaw relationship of volatility and momentum stocks*. Hentet fra blackrockblog: https://www.blackrockblog.com/2017/03/30/seesaw-volatility-momentum/

Kothari, S. P., Shanken, J., & Richard, S. G. (1995). Another look at the cross-section of expected stock returns. *The journal of finance 50(1)*, 185-224.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *The journal of finance, 54(2)*, 471-517.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (2002). Investor protection and corporate valuation. *The journal of finance, 57(3)*, 1147-1170.

Lakonishok, J., & Shapiro, A. C. (1986). Systematic risk, total risk and size as determinants of stock market returns. *Journal of Banking & Finance, 10(1*, 115-132.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The journal of finance, 49(5)*, 1541-1578.

Lau, K., Moe, T., Bei, B., & Liu, C. (2014). *SH-HK Connect: New regime, unprecedented opportunity - The 'new' investment case for China for global/A-share investors.* Beijing: Goldman Sachs. Hentet 15. September 2018 fra https://www.msci.com/eqb/pressreleases/archive/2017_Market_Classification_Announcement_Press_Release_FINAL.pdf

Lin, Q. (2017). Noisy prices and the Fama–French five-factor asset pricing model in China. *Emerging Markets Review, 31* , 141-163.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance, 20(4)*, 587-615.

Liu, L. X., & Zhang, L. (2008). Momentum profits, factor pricing and macroeconomic risk. *The Review of Financial Studies, 21(6)*, 2417-2448.

logic, A. p. (2007). Eun, C. S.; Huang, W. *Pacific-Basin Finance Journal, 15(5)*, 452-480.

Ma, C. (10. November 2014). *Understanding Stock Connect*. Hentet 14. 09 2016 fra Goldman Sachs: https://www.goldmansachs.com/insights/pages/stock-connect-video.html

MacKinlay, A. C. (1987). On multivariate tests of the CAPM. *Journal of Financial Economics, 18(2)*, 341-371.

MacKinlay, A. C. (1995). Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics, 38(1)*, 3-28.

Markowitz, H. (1952). Portfolio selection. *he journal of finance, 7(1)*, 77-91.

McConnell, J. J., & & Muscarella, C. J. (1985). Corporate capital expenditure decisions and the market value of the firm. *Journal of financial economics, 14(3)*, 399-422.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867-887.

Miller, M. H., & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business, 34(4)*, 411-433.

MSCI. (2017). *RESULTS OF MSCI 2017 MARKET CLASSIFICATION REVIEW.* MSCI. Hentet 15. September 2018 fra MSCI: https://www.msci.com/eqb/pressreleases/archive/2017_Market_Classification_Announcement_Press_Release_FINAL.pdf

MSCI. (4. May 2018). *MSCI* . Hentet fra MSCI constituents: https://www.msci.com/constituents

MSCI ANNOUNCES THE RESULTS OF ITS ANNUAL MARKET CLASSIFICATION REVIEW. (u.d.). Hentet 15. September 2018 fra MSCI: https://www.msci.com/market-classification

MSCI Global. (2018). *MSCI GLOBAL INVESTABLE MARKET INDEXES METHODOLOGY.* MSCI.

Munk, C. (2017). *Financial markets and investments.* Lecture Notes, Copenhagen Business School.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108(1)*, 1-28.

On the predictability of Chinese stock returns. (2010). *Pacific-Basin Finance Journal 18*, 403-425.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of financial Economic, 4(2)*, 129-176.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Efficient Capital Markets: II. *Persuasive Evidence of Market Inefficiency, 11(3)*, 9-16.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory, 13*, 341-360.

Sagi, J. S. (2007). Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics, 84(2)*, 389-434.

Schmidt, P., Von Arx, U., Schrimpf, A., Wagner, A., & Ziegler, A. (2017). On the construction of common size, value and momentum factors in international stock markets: A guide with applications. *Economics Working Paper Series, No. 11/141*.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19(3)*, 425-442.

Stattman, D. (1980). Book values and stock returns. *The Chicago MBA: A journal of selected papers, 4(1)*, 25-45.

Titman, S., Wei, K. J., & Xie, F. X. (2001). Explaining the cross-section of stock returns in Japan: Factors or characteristics? *The Journal of Finance, 56(2)*, 743-766.

Titman, S., Wei, K. J., & Xie, F. X. (2004). Capital investments and stock returns. *Journal of financial and Quantitative Analysis, 39(4)*, 677-700.

Wang, F., & Xu, Y. (2004). What Determines Chinese Stock Returns? *Financial Analysts Journals, 60(6)*, 65–77.

WHAT IS STOCK CONNECT. (u.d.). Hentet 15. September 2018 fra hkex: http://www.hkex.com.hk/Mutual-Market/Stock-Connect?sc_lang=en

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach.* Mason, OH 45040: South-Western Cengage Learning.

World Bank. (03. January 2018). *World Development Indicators*. Hentet fra databank.worldbank: http://databank.worldbank.org/data/reports.aspx?source=2&series=CM.MKT.TRAD.GD.ZS

Xie, S., & Qu, Q. (2016). The three-factor model and size and value premiums in china's stock market. *Emerging Markets Finance and Trade, 52(5)*, 1092-1105.

Xu, J., & Zhang, S. (2014). The Fama-French Three Factors in the Chinese Stock Market. *China Accounting and Finance Review*.

Appendix

Appendix 1 – Numbers GDP development over time

Year	China	Japan	Germany	United States
1987	272,973	2,514,284	1,293,264	4,870,217
1988	312,354	3,050,638	1,395,932	5,252,629
1989	347,768	3,052,316	1,393,674	5,657,693
1990	360,858	3,139,974	1,764,968	5,979,589
1991	383,373	3,578,139	1,861,874	6,174,043
1992	426,916	3,897,826	2,123,131	6,539,299
1993	444,731	4,466,565	2,068,556	6,878,718
1994	564,325	4,907,039	2,205,966	7,308,755
1995	734,548	5,449,116	2,591,620	7,664,060
1996	863,747	4,833,713	2,503,665	8,100,201
1997	961,604	4,414,733	2,218,689	8,608,515
1998	1,029,043	4,032,510	2,243,226	9,089,168
1999	1,093,997	4,562,079	2,199,957	9,660,624
2000	1,211,347	4,887,520	1,949,954	10,284,779
2001	1,339,396	4,303,544	1,950,649	10,621,824
2002	1,470,550	4,115,116	2,079,136	10,977,514
2003	1,660,288	4,445,658	2,505,734	11,510,670
2004	1,955,347	4,815,149	2,819,245	12,274,928
2005	2,285,966	4,755,411	2,861,410	13,093,726
2006	2,752,132	4,530,377	3,002,446	13,855,888
2007	3,552,182	4,515,265	3,439,953	14,477,635
2008	4,598,206	5,037,908	3,752,366	14,718,582
2009	5,109,954	5,231,383	3,418,005	14,418,739
2010	6,100,620	5,700,098	3,417,095	14,964,372
2011	7,572,554	6,157,460	3,757,698	15,517,926
2012	8,560,547	6,203,213	3,543,984	16,155,255
2013	9,607,224	5,155,717	3,752,514	16,691,517
2014	10,482,372	4,848,733	3,890,607	17,393,103
2015	11,064,666	4,383,076	3,375,611	18,120,714
2016	11,199,145	4,940,159	3,477,796	18,624,475

Appendix 2 – Annual GDP growth rates over time (in percentage)

Year	China	Euro area	Japan	Germany	US	UK	East Asia	World
1987	11.69	2.55	4.11	1.40	3.46	5.31	5.50	3.58
1988	11.23	4.31	7.15	3.71	4.20	5.75	7.66	4.65
1989	4.19	4.10	5.37	3.90	3.68	2.57	5.42	3.75
1990	3.91	3.58	5.57	5.26	1.92	0.73	5.51	3.00
1991	9.29	2.66	3.32	5.11	-0.07	-1.09	4.40	1.43
1992	14.22	1.41	0.82	1.92	3.56	0.37	3.42	1.79
1993	13.87	-0.66	0.17	-0.96	2.75	2.53	3.58	1.63
1994	13.05	2.47	0.86	2.46	4.04	3.88	4.27	3.01
1995	10.95	2.47	2.74	1.74	2.72	2.47	5.11	3.05
1996	9.93	1.72	3.10	0.82	3.80	2.54	5.09	3.38
1997	9.23	2.72	1.08	1.85	4.49	4.04	3.46	3.71
1998	7.84	2.97	-1.13	1.98	4.45	3.14	-0.11	2.52
1999	7.67	2.99	-0.25	1.99	4.69	3.22	2.99	3.26
2000	8.49	3.86	2.78	2.96	4.09	3.66	4.88	4.40
2001	8.34	2.16	0.41	1.70	0.98	2.54	2.67	1.92
2002	9.13	1.02	0.12	0.00	1.79	2.46	3.59	2.15
2003	10.04	0.72	1.53	-0.71	2.81	3.33	4.26	2.91
2004	10.11	2.32	2.20	1.17	3.79	2.36	5.15	4.46
2005	11.40	1.70	1.66	0.71	3.35	3.10	5.05	3.84
2006	12.72	3.23	1.42	3.70	2.67	2.46	5.55	4.32
2007	14.23	3.05	1.65	3.26	1.78	2.36	6.50	4.25
2008	9.65	0.44	-1.09	1.08	-0.29	-0.47	3.50	1.82
2009	9.40	-4.52	-5.42	-5.62	-2.78	-4.19	1.35	-1.74
2010	10.64	2.09	4.19	4.08	2.53	1.69	7.06	4.32
2011	9.54	1.60	-0.12	3.66	1.60	1.45	4.60	3.18
2012	7.86	-0.89	1.50	0.49	2.22	1.48	4.66	2.45
2013	7.76	-0.24	2.00	0.49	1.68	2.05	4.75	2.63
2014	7.30	1.33	0.34	1.93	2.57	3.05	4.12	2.85
2015	6.90	2.08	1.22	1.74	2.86	2.35	4.11	2.82
2016	6.69	1.81	1.03	1.94	1.49	1.79	4.11	2.49

Appendix 3 – Overview of Stata Do-File

Do-File name	**Purpose**
General Do Files used to perform the analysis	
Data sorting.do	- Load the data into Stata needed for three-factor model - clean the data - reshape it to a time series format
Data Screening.do	- Static and dynamic screening of the data
Creation_PortfolioV01.do	- Create the market return of the entire stock market - the sorting variables and specific portfolio identifiers - perform general summary statistics
Five Factor inputs.do	- Merge data in used for five factor model performance into the data set - Merge risk free rate into the data set, generate identifiers used for portfolio sorts of five factor model - Create dependent variable portfolios (5x5 sorts and 2x4x4 sorts), their summary statistics and excess return portfolios - Exclusion of financial firms
IndependentVariableV01.do	- Create Factors for the three-factor model - Create momentum factors
Indep_5FF new.do	- Create Factors for the five factor model for Size-B/M and Size-Momentum sorts - Create average factor premia statistics for several time periods
Indep_5FF new for different Dep Var Portfolio Sortings.do	- Create Factors for the five factor model for alternative sorts (Size-OP and Size-Inv) - Create average factor premia statistics for several time periods
Regression Analysis.do	- Perform CAPM, two-factor model, three-factor model, four factor model and five factor model regressions for all four sorts
GRS test.do	- Reshape the data and performing the GRS test for the models for Size-B/M and Size-Momentum sorts
GRS test alternative sorts.do	- Reshape the data and performing the GRS test for the models for alternative sorts (Size-OP and Size-Inv)

Do files used for additional graphs	
Break_up_Risk_premia.do	- Closer investigation of the Risk premia
Summary Stats Period 2006-2017.do	- Summary statistics over the observation period 2006-2017
Univariate_Sorting.do	- Univariate sorts of the variables of interest (Size, B/M, Momentum, OP and Inv)
Do files used for Robustness checks	
Total_MV_Yearly_Average_Robustness.do	- Replace the Total_MV investigated at one point in time with the average Total_MV over the last year
Robustness_Switched_Periods.do	- Switch the period of investigation for the GRS tests on the three-factor model
Robustness_Shorter_Momentum.do	- Create shorter momentum periods (three-month and two-month period)
Robustness_Orthogonal_HML.do	- Return the three factor model with a orthogonal HML factor
Robustness_Equally_weighted_RiskFactor.do	- Create equally weighted instead of value weighted portfolios
Robustness_Different_book_calculation.do	- Change the B/M calculation from using Total_MV to using Float_MV
Robustness_Changed_RF.do	- Changing the Risk Free rate from a bank deposit rate into a TBILL rate
RM_MSCI_Index_Robustness.do	- Use the MSCI China Index as a market return rate
Float_Share_Size_Factor_Robustness_Test.do	- change the Size Factor from the Total_MV to Float_MV
Do files used to manipulate data before usage	
Merge Rf.do	- Merge the risk free rate
MSCI.do	- Used to load and manipulate the MSCI market index
RF_NEW_China_TBILL.do	- Load and manipulate the Chinese t-bill rate

Appendix 4 - Excess Returns

Excess Returns

Sort on Size - B/M (5x5)

Excess Returns

	Low	2	3	4	High
Small	2.46%	2.47%	2.42%	2.16%	2.23%
2	2.16%	2.27%	1.85%	1.64%	1.67%
3	2.00%	2.00%	1.59%	1.38%	1.27%
4	2.18%	1.36%	1.21%	1.22%	1.17%
Big	1.57%	1.26%	0.81%	0.52%	0.95%

Standard Deviation

	Low	2	3	4	High
Small	3.15%	3.53%	3.38%	3.48%	3.41%
2	3.47%	3.51%	3.30%	3.33%	3.42%
3	3.76%	3.12%	3.28%	3.31%	3.74%
4	3.58%	3.39%	3.58%	3.75%	3.94%
Big	2.99%	3.84%	3.65%	3.65%	3.70%

t-values

	Low	2	3	4	High
Small	2.587	2.316	2.375	2.058	2.162
2	2.064	2.138	1.855	1.629	1.619
3	1.765	2.124	1.613	1.380	1.123
4	2.014	1.333	1.122	1.080	0.983
Big	1.743	1.089	0.741	0.473	0.855

Sort on Size - Momentum (5x5)

Excess Returns

	Looser	2	3	4	Winner
Small	4.92%	2.83%	3.37%	2.53%	2.97%
2	4.74%	2.73%	2.67%	2.10%	1.88%
3	3.04%	2.45%	1.99%	2.00%	1.40%
4	3.06%	2.22%	2.25%	1.70%	1.77%
Big	2.75%	2.28%	1.27%	1.50%	1.16%

Standard Deviation

	Looser	2	3	4	Winner
Small	3.77%	3.44%	3.43%	3.43%	3.30%
2	3.34%	3.80%	3.33%	3.36%	3.09%
3	3.32%	3.50%	3.47%	3.66%	3.15%
4	3.58%	3.86%	3.35%	3.63%	3.53%
Big	4.54%	3.37%	3.79%	3.16%	3.35%

t-values

	Looser	2	3	4	Winner
Small	2.587	2.316	2.375	2.058	2.162
2	2.064	2.138	1.855	1.629	1.619
3	1.765	2.124	1.613	1.380	1.123
4	2.014	1.333	1.122	1.080	0.983
Big	1.743	1.089	0.741	0.473	0.855

Sort on Size - OP (5x5)

Excess Returns

	Low	2	3	4	High
Small	2.42%	2.29%	2.41%	2.22%	1.13%
2	1.82%	1.60%	1.79%	1.11%	2.14%
3	1.45%	2.10%	2.25%	1.01%	0.92%
4	1.46%	1.70%	1.45%	1.23%	0.96%
Big	1.10%	1.12%	0.38%	1.17%	0.53%

Standard Deviation

	Low	2	3	4	High
Small	3.33%	3.54%	3.47%	3.28%	1.77%
2	3.50%	3.24%	3.45%	2.57%	3.62%
3	3.70%	3.93%	3.68%	2.46%	2.67%
4	3.55%	3.58%	3.55%	3.26%	3.20%
Big	4.50%	4.18%	2.92%	3.49%	3.24%

t-values

	Low	2	3	4	High
Small	2.411	2.148	2.302	2.146	1.691
2	1.726	1.713	1.725	1.361	1.566
3	1.302	1.511	2.117	1.301	1.095
4	1.364	1.647	1.353	1.250	0.850
Big	0.772	0.848	0.416	1.115	0.486

Sort on Size - Inv (5x5)

Excess Returns

	Low	2	3	4	High
Small	2.40%	2.25%	2.76%	2.00%	2.78%
2	2.09%	1.25%	2.02%	1.93%	0.89%
3	1.62%	2.41%	1.14%	1.64%	1.36%
4	1.53%	0.79%	1.18%	1.35%	0.74%
Big	1.79%	1.07%	0.56%	0.36%	0.40%

Standard Deviation

	Low	2	3	4	High
Small	3.37%	3.23%	3.54%	3.31%	3.89%
2	3.65%	2.87%	3.31%	3.46%	2.75%
3	3.75%	3.73%	2.83%	3.13%	3.34%
4	4.17%	2.90%	3.00%	3.35%	2.94%
Big	4.21%	3.45%	3.10%	2.97%	3.56%

t-values

	Low	2	3	4	High
Small	2.356	2.307	2.467	2.002	2.148
2	1.809	1.379	1.936	1.761	1.076
3	1.364	2.148	1.279	1.735	1.353
4	1.161	0.867	1.178	1.341	0.800
Big	1.344	1.025	0.573	0.386	0.320

Appendix 5 – Excess returns for advanced portfolio sorts

2x4x4 Size-B/M-OP									
	Small					Big			
	B/M - OP Excess Returns					B/M - OP Excess Returns			
	Low	**2**	**3**	**High**		**Low**	**2**	**3**	**High**
Small	0.60%	1.04%	2.53%	0.54%	**Small**	-1.10%	0.50%	1.48%	1.32%
2	2.75%	2.02%	1.50%	2.57%	**2**	1.53%	-1.10%	-0.01%	1.80%
3	3.33%	1.14%	2.83%	1.47%	**3**	2.80%	3.24%	-0.69%	0.32%
Big	1.83%	2.37%	-0.41%	1.25%	**Big**	1.65%	1.20%	2.11%	1.20%

2x4x4 Size-B/M-Inv									
	Small					Big			
	B/M - Inv Excess Returns					B/M - Inv Excess Returns			
	Low	**2**	**3**	**High**		**Low**	**2**	**3**	**High**
Small	-0.20%	1.97%	0.91%	2.25%	**Small**	2.13%	1.47%	1.36%	-0.96%
2	3.43%	1.81%	2.50%	2.20%	**2**	0.31%	2.30%	-1.94%	2.60%
3	2.15%	-1.35%	3.84%	1.31%	**3**	1.22%	-0.01%	1.35%	0.10%
Big	5.22%	2.43%	0.63%	2.39%	**Big**	1.08%	2.26%	-0.41%	2.37%

Appendix 6 – Regressions on Size - B/M

Sort on Size - B/M (5x5)

4-Factor Model

Market ß	Low	2	3	4	High
Small	0.93064	1.00686	1.01934	1.00945	0.98521
2	0.98913	1.06948	1.02428	1.02439	1.02858
3	1.04082	1.05422	1.04987	1.04814	1.02996
4	1.02347	1.02066	1.04085	1.08353	1.0417
Big	0.99467	1.05861	1.06534	0.99765	0.96325

t(ß)	Low	2	3	4	High
Small	20.1914	22.6121	29.8636	29.073	31.2449
2	22.7838	26.1236	34.0221	29.7251	33.8271
3	22.638	26.4005	27.2036	30.7614	28.9232
4	29.1394	24.0219	22.834	26.0711	29.6011
Big	33.3383	24.0631	26.8612	22.3403	28.6105

Factor loading SMB (s)	Low	2	3	4	High
Small	1.08557	1.39874	1.28296	1.40202	1.37427
2	1.28922	1.13845	1.18523	1.21422	1.18493
3	0.96219	0.98769	1.00412	1.0286	1.07445
4	0.81406	0.86204	0.68895	0.69955	0.671
Big	0.06963	0.19762	-0.0267	0.02176	-0.1798

t(s)	Low	2	3	4	High
Small	12.1815	13.7036	17.9603	17.761	18.9716
2	14.2328	11.3331	17.6019	15.407	17.7735
3	10.0319	14.1516	11.409	12.3838	11.5564
4	9.55399	9.1616	6.7786	5.80842	7.64074
Big	1.11273	1.6223	-0.2732	0.18778	-2.3148

Factor loading HML (h)	Low	2	3	4	High
Small	-0.4048	-0.1396	0.00097	0.19103	0.35127
2	-0.1285	-0.254	-0.0412	0.19329	0.50607
3	-0.3032	-0.2732	-0.0861	0.18611	0.59568
4	-0.6	-0.2324	-0.1019	0.23315	0.61545
Big	-0.6257	-0.1977	-0.0272	0.27448	0.80736

t(h)	Low	2	3	4	High
Small	-3.20	-1.22	0.01	2.68	6.01
2	-1.61	-3.05	-0.65	2.66	8.12
3	-3.28	-3.52	-1.15	2.08	8.11
4	-5.94	-2.51	-0.94	1.76	7.84
Big	-12.30	-1.67	-0.26	2.37	9.38

Factor loading WML (w)	Low	2	3	4	High
Small	-0.3194	0.05794	0.07019	0.06796	0.07967
2	0.1236	0.17688	0.17066	0.13611	0.02744
3	0.17917	0.14139	0.06656	0.11048	0.12565
4	0.12304	0.22104	0.13284	0.14055	-0.0253
Big	0.14359	0.02579	-0.0995	0.01386	0.09157

t(w)	Low	2	3	4	High
Small	-2.93	0.58	1.13	1.01	1.28
2	1.42	2.42	2.43	1.90	0.48
3	1.94	1.77	0.79	1.35	1.65
4	1.19	2.25	1.35	1.29	-0.31
Big	2.48	0.21	-1.07	0.11	1.01

R^2	Low	2	3	4	High
Small	0.85	0.93256	0.94362	0.94828	0.94879
2	0.9077	0.93896	0.94745	0.93685	0.9539
3	0.91099	0.92757	0.9289	0.93773	0.93694
4	0.9151	0.9165	0.88928	0.90558	0.92829
Big	0.92855	0.88899	0.90404	0.87173	0.92585

α	Low	2	3	4	High
Small	0.17%	0.25%	0.37%	0.12%	0.33%
2	0.01%	0.23%	-0.07%	-0.20%	-0.06%
3	0.13%	0.08%	-0.34%	-0.35%	-0.25%
4	0.24%	-0.35%	-0.35%	-0.21%	-0.11%
Big	0.25%	-0.02%	-0.28%	-0.37%	0.67%

t(α)	Low	2	3	4	High
Small	0.48703	0.88745	1.53772	0.48981	1.40228
2	0.04537	0.85173	-0.3059	-0.8078	-0.278
3	0.41367	0.2931	-1.2803	-1.4077	-1.054
4	0.85277	-1.2773	-1.1511	-0.7684	-0.4509
Big	1.13605	-0.0829	-1.0106	-1.2108	2.76155

5-Factor Model

Sort on Size - B/M (5x5)

Market ß	Low	2	3	4	High
Small	0.95113	0.98518	1.00313	0.96958	0.95233
2	0.95276	1.01143	0.96663	0.97554	1.00986
3	0.9614	1.01045	0.99611	1.00713	0.98915
4	0.98454	0.98209	0.99523	1.0374	1.02342
Big	0.96351	1.0031	1.01932	1.00501	0.92508

t(ß)	Low	2	3	4	High
Small	15.4765	16.3745	22.5154	20.7809	21.1687
2	16.7072	21.982	24.5426	22.9284	23.0664
3	16.9802	20.6809	20.8372	23.4942	19.9518
4	23.0993	18.0027	17.443	21.687	20.5788
Big	22.8552	18.3585	22.0777	16.1856	20.3447

Factor loading SMB (s)	Low	2	3	4	High
Small	1.07224	1.17434	1.11035	1.24294	1.2217
2	1.11881	1.01879	0.963	1.04759	1.14227
3	0.60042	0.79357	0.83437	0.86998	0.91641
4	0.58494	0.64345	0.45076	0.50039	0.55159
Big	-0.0423	0.08975	-0.1322	-0.186	-0.3489

t(s)	Low	2	3	4	High
Small	7.7499	10.2837	12.7582	12.6866	13.1491
2	10.0556	12.2445	12.5809	11.179	13.365
3	6.1448	8.21199	7.72507	9.75353	9.49743
4	6.38537	6.37274	4.06595	4.24428	5.72421
Big	-0.5027	0.79032	-1.2336	-1.6541	-4.0358

Factor loading HML (h)	Low	2	3	4	High
Small	-0.3935	-0.2417	-0.0573	0.17896	0.33258
2	-0.1868	-0.2169	-0.044	0.18953	0.49852
3	-0.3075	-0.2988	-0.0439	0.17542	0.57497
4	-0.6187	-0.2911	-0.1294	0.21135	0.61446
Big	-0.6769	-0.1104	0.08085	0.17979	0.79172

t(h)	Low	2	3	4	High
Small	-2.87	-1.96	-0.67	2.12	4.36
2	-1.77	-3.20	-0.57	2.41	6.12
3	-3.36	-3.78	-0.55	2.10	6.18
4	-7.67	-3.29	-1.17	1.70	6.23
Big	-8.64	-0.98	0.69	1.44	9.62

Factor loading RMW (r)	Low	2	3	4	High
Small	0.07899	-0.3634	-0.298	-0.3463	-0.3341
2	-0.2753	-0.4567	-0.5112	-0.3905	-0.1084
3	-0.8062	-0.4233	-0.4373	-0.3728	-0.3481
4	-0.4739	-0.4601	-0.5202	-0.4133	-0.2253
Big	-0.2119	-0.372	-0.3388	-0.2895	-0.3481

t(r)	Low	2	3	4	High
Small	0.42	-2.32	-2.14	-2.23	-2.05
2	-1.59	-3.83	-3.50	-2.47	-0.79
3	-5.35	-2.91	-2.70	-2.32	-1.94
4	-2.54	-2.30	-2.82	-2.23	-1.29
Big	-1.32	-2.11	-1.67	-1.53	-2.12

Factor loading CMA (c)	Low	2	3	4	High
Small	0.45216	0.1351	0.03119	-0.1839	-0.1587
2	-0.0833	-0.377	-0.3908	-0.3979	-0.1372
3	-0.4211	-0.289	-0.4502	-0.2415	-0.2453
4	-0.1575	-0.2608	-0.2717	-0.2578	-0.0027
Big	-0.2767	-0.4888	-0.3585	0.24868	-0.2085

t(c)	Low	2	3	4	High
Small	1.63	0.58	0.16	-1.00	-0.91
2	-0.29	-2.06	-2.46	-2.31	-0.80
3	-1.72	-1.45	-2.39	-1.33	-1.21
4	-0.82	-1.28	-1.18	-1.23	-0.01
Big	-1.54	-2.06	-1.53	1.01	-0.98

R^2	Low	2	3	4	High
Small	0.85	0.93256	0.94362	0.94828	0.94879
2	0.9077	0.93896	0.94745	0.93685	0.9539
3	0.91099	0.92757	0.9289	0.93773	0.93694
4	0.9151	0.9165	0.88928	0.90558	0.92829
Big	0.92855	0.88899	0.90404	0.87173	0.92585

α	Low	2	3	4	High
Small	0.33%	0.26%	0.36%	0.19%	0.36%
2	0.10%	0.12%	-0.02%	-0.13%	-0.02%
3	0.27%	0.12%	-0.14%	-0.33%	-0.22%
4	0.24%	-0.38%	-0.28%	-0.17%	-0.01%
Big	0.29%	0.13%	-0.03%	-0.30%	0.73%

t(α)	Low	2	3	4	High
Small	0.89422	0.98201	1.47886	0.78847	1.51495
2	0.31879	0.59459	-0.1044	-0.5609	-0.0928
3	1.01792	0.43977	-0.5872	-1.3114	-0.8838
4	0.89785	-1.4375	-0.9167	-0.5995	-0.0551
Big	1.2334	0.46253	-0.0876	-1.0074	2.86582

Appendix 7 – Regressions on Size – Momentum sort

3-Factor Model

Sort on Size - Momentum (5x5)

Market ß	Low	2	3	4	High
Small	0.97	0.99	1.00	1.03	1.00
2	1.00	1.04	1.02	1.05	1.07
3	1.01	1.03	1.04	1.08	1.09
4	1.01	1.06	1.02	1.07	1.06
Big	1.02	0.96	1.04	0.97	1.02

t(ß)	Low	2	3	4	High
Small	22.699	29.719	30.191	29.548	17.421
2	26.478	32.957	29.625	28.989	29.107
3	25.117	26.785	27.219	30.684	25.873
4	20.619	27.624	29.497	24.620	24.995
Big	20.981	16.126	26.120	22.305	20.150

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.38	1.43	1.29	1.25	1.27
2	1.23	1.29	1.17	1.13	0.97
3	1.12	1.08	0.99	0.97	0.82
4	0.96	0.81	0.71	0.65	0.53
Big	0.19	-0.04	0.08	-0.11	-0.03

t(s)	Low	2.000	3.000	4.000	High
Small	16.035	19.521	17.647	14.241	10.534
2	16.023	17.039	16.004	14.792	12.465
3	13.038	13.159	11.353	11.876	7.656
4	9.675	9.333	7.713	6.498	5.367
Big	1.715	-0.205	0.885	-1.205	-0.268

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	0.15	0.23	0.02	0.09	-0.10
2	0.18	0.27	0.14	0.01	-0.27
3	0.21	0.15	0.01	0.00	-0.24
4	0.20	0.14	-0.01	-0.28	-0.41
Big	0.21	0.22	0.12	-0.05	-0.30

t(h)	Low	2.000	3.000	4.000	High
Small	1.414	4.388	0.413	0.994	-0.915
2	3.766	4.491	2.015	0.201	-3.944
3	3.779	2.338	0.236	0.057	-2.290
4	2.453	1.927	-0.092	-3.416	-3.678
Big	1.956	1.660	1.569	-0.653	-2.264

R^2	Low	2.00	3.00	4.00	High
Small	0.92	0.95	0.94	0.94	0.88
2	0.93	0.95	0.94	0.93	0.93
3	0.93	0.93	0.93	0.93	0.89
4	0.89	0.91	0.92	0.89	0.88
Big	0.84	0.81	0.89	0.89	0.83

α	Low	2.00	3.00	4.00	High
Small	1.15%	0.52%	0.56%	-0.11%	-1.07%
2	-0.06%	0.37%	0.15%	-0.41%	-1.12%
3	-0.02%	0.09%	-0.04%	-0.43%	-0.85%
4	-0.22%	0.01%	-0.01%	-0.27%	-0.80%
Big	-0.04%	0.13%	0.11%	0.07%	-0.11%

t(α)	Low	2.000	3.000	4.000	High
Small	3.787	2.164	2.391	-0.440	-2.949
2	-0.252	1.584	0.625	-1.597	-4.306
3	-0.071	0.356	-0.142	-1.649	-2.680
4	-0.674	0.030	-0.020	-0.878	-2.532
Big	-0.112	0.335	0.379	0.255	-0.303

4-Factor Model

Sort on Size - Momentum (5x5)

Market ß

	Low	2	3	4	High
Small	0.98	1.00	1.00	1.02	0.99
2	1.00	1.04	1.01	1.04	1.05
3	1.02	1.04	1.03	1.07	1.07
4	1.02	1.06	1.02	1.06	1.04
Big	1.04	0.97	1.03	0.96	0.99

t(ß)

	Low	2	3	4	High
Small	24.378	30.162	29.922	30.140	17.234
2	27.995	33.062	28.693	28.541	29.590
3	27.605	27.707	26.573	30.009	26.680
4	22.378	27.521	28.939	24.313	26.350
Big	24.780	17.393	25.755	20.959	26.991

Factor loading SMB (s)

	Low	2.00	3.00	4.00	High
Small	1.32	1.41	1.29	1.30	1.33
2	1.20	1.28	1.20	1.17	1.04
3	1.07	1.06	1.02	1.00	0.91
4	0.90	0.80	0.73	0.70	0.63
Big	0.10	-0.11	0.10	-0.07	0.12

t(s)

	Low	2.000	3.000	4.000	High
Small	13.319	19.379	17.357	16.261	11.397
2	15.045	16.399	17.302	16.233	16.626
3	12.369	12.726	12.532	13.189	10.298
4	7.632	8.273	7.412	7.139	7.497
Big	0.724	-0.738	1.138	-0.735	1.429

Factor loading HML (h)

	Low	2.00	3.00	4.00	High
Small	0.00	0.18	0.02	0.22	0.07
2	0.09	0.25	0.21	0.12	-0.07
3	0.08	0.08	0.09	0.10	0.00
4	0.05	0.10	0.03	-0.16	-0.14
Big	-0.04	0.02	0.17	0.06	0.07

t(h)

	Low	2.000	3.000	4.000	High
Small	0.012	2.971	0.264	2.836	0.655
2	1.452	3.160	2.836	1.694	-1.303
3	1.057	1.078	1.357	1.398	-0.023
4	0.393	0.955	0.298	-1.673	-1.498
Big	-0.223	0.106	2.366	0.664	0.781

Factor loading WML (w)

	Low	2.00	3.00	4.00	High
Small	-0.36	-0.11	-0.01	0.33	0.42
2	-0.20	-0.06	0.17	0.26	0.47
3	-0.30	-0.16	0.19	0.23	0.55
4	-0.36	-0.10	0.09	0.28	0.63
Big	-0.58	-0.48	0.13	0.26	0.89

t(w)

	Low	2.000	3.000	4.000	High
Small	-3.185	-1.757	-0.192	4.419	3.258
2	-2.589	-0.776	2.613	3.642	7.636
3	-3.535	-1.977	2.738	2.971	6.256
4	-2.672	-0.899	0.883	3.036	8.431
Big	-3.769	-2.649	1.523	2.835	11.303

R^2

	Low	2.00	3.00	4.00	High
Small	0.93	0.95	0.94	0.95	0.89
2	0.94	0.95	0.94	0.94	0.94
3	0.94	0.93	0.93	0.94	0.92
4	0.90	0.91	0.92	0.90	0.92
Big	0.88	0.83	0.89	0.89	0.91

α

	Low	2	3	4	High
Small	0.85%	0.43%	0.55%	0.15%	-0.73%
2	-0.23%	0.32%	0.29%	-0.20%	-0.73%
3	-0.26%	-0.03%	0.12%	-0.24%	-0.40%
4	-0.52%	-0.07%	0.07%	-0.04%	-0.29%
Big	-0.51%	-0.25%	0.21%	0.29%	0.61%

t(α)

	Low	2	3	4	High
Small	3.234	1.776	2.369	0.671	-1.871
2	-0.931	1.307	1.231	-0.790	-3.156
3	-1.103	-0.123	0.450	-0.971	-1.458
4	-1.690	-0.253	0.259	-0.128	-1.121
Big	-1.847	-0.718	0.736	1.009	2.323

5-Factor Model

Sort on Size - Momentum (5x5)

Market ß

	Low	2	3	4	High
Small	0.96	0.98	0.99	1.00	0.91
2	0.96	0.99	0.96	1.00	1.03
3	0.97	0.98	0.97	1.01	1.03
4	0.95	1.01	0.97	1.03	1.03
Big	1.00	0.95	0.99	0.94	0.98

t(ß)

	Low	2	3	4	High
Small	17.254	22.387	22.044	22.260	13.147
2	19.936	25.857	22.127	21.980	20.347
3	19.604	21.219	20.876	22.144	18.421
4	15.641	21.886	22.053	17.548	19.659
Big	17.063	12.158	19.722	18.167	15.100

Factor loading SMB (s)

	Low	2.00	3.00	4.00	High
Small	1.33	1.40	1.22	1.10	0.90
2	1.18	1.20	1.02	1.01	0.85
3	1.07	0.97	0.84	0.80	0.57
4	0.76	0.70	0.58	0.52	0.34
Big	0.12	-0.04	-0.05	-0.18	-0.25

t(s)

	Low	2.000	3.000	4.000	High
Small	14.304	12.887	14.743	11.275	6.415
2	11.810	13.934	12.799	10.681	8.923
3	10.814	9.826	9.099	9.114	4.553
4	5.813	6.569	5.889	4.993	3.276
Big	0.957	-0.258	-0.548	-1.676	-1.911

Factor loading HML (h)

	Low	2.00	3.00	4.00	High
Small	0.16	0.26	0.02	0.10	-0.02
2	0.26	0.36	0.24	0.08	-0.22
3	0.32	0.25	0.13	0.11	-0.18
4	0.30	0.20	0.08	-0.21	-0.40
Big	0.24	0.25	0.20	0.01	-0.29

t(h)

	Low	2.000	3.000	4.000	High
Small	1.191	3.362	0.290	1.090	-0.130
2	3.640	5.014	2.873	0.872	-2.656
3	3.858	2.981	1.735	1.330	-1.623
4	2.743	2.167	0.841	-2.001	-3.521
Big	1.653	1.425	2.306	0.094	-1.853

Factor loading RMW (r)

	Low	2.00	3.00	4.00	High
Small	-0.10	-0.11	-0.15	-0.34	-0.91
2	-0.21	-0.33	-0.47	-0.34	-0.31
3	-0.26	-0.39	-0.49	-0.51	-0.63
4	-0.56	-0.34	-0.41	-0.37	-0.42
Big	-0.19	-0.06	-0.40	-0.23	-0.48

t(r)

	Low	2.000	3.000	4.000	High
Small	-0.558	-0.653	-0.993	-2.224	-3.926
2	-1.252	-2.204	-3.432	-2.285	-1.967
3	-1.694	-2.189	-3.432	-3.166	-2.644
4	-2.474	-1.581	-2.101	-1.825	-2.080
Big	-0.899	-0.203	-2.296	-1.341	-1.971

Factor loading CMA (c)

	Low	2.00	3.00	4.00	High
Small	-0.01	-0.12	0.01	-0.04	-0.27
2	-0.30	-0.36	-0.39	-0.24	-0.17
3	-0.45	-0.38	-0.44	-0.40	-0.16
4	-0.36	-0.24	-0.33	-0.24	0.02
Big	-0.11	-0.15	-0.32	-0.24	0.03

t(c)

	Low	2.000	3.000	4.000	High
Small	-0.038	-0.678	0.063	-0.251	-0.882
2	-1.478	-2.339	-2.323	-1.223	-0.820
3	-2.085	-2.064	-2.326	-2.212	-0.692
4	-1.269	-1.183	-1.847	-1.030	0.083
Big	-0.386	-0.435	-1.636	-0.950	0.086

R^2

	Low	2.00	3.00	4.00	High
Small	0.91	0.95	0.94	0.94	0.89
2	0.94	0.95	0.94	0.93	0.93
3	0.93	0.93	0.93	0.94	0.90
4	0.89	0.91	0.92	0.90	0.89
Big	0.84	0.81	0.89	0.89	0.83

α

	Low	2	3	4	High
Small	1.17%	0.58%	0.59%	-0.03%	-0.82%
2	0.05%	0.52%	0.33%	-0.29%	-1.01%
3	0.14%	0.26%	0.17%	-0.23%	-0.69%
4	-0.02%	0.13%	0.15%	-0.14%	-0.72%
Big	0.02%	0.18%	0.27%	0.18%	-0.02%

t(α)

	Low	2	3	4	High
Small	3.888	2.290	2.567	-0.132	-2.471
2	0.198	2.367	1.658	-1.160	-3.841
3	0.543	0.982	0.726	-0.909	-2.213
4	-0.074	0.451	0.602	-0.452	-2.302
Big	0.067	0.474	0.918	0.656	-0.060

Appendix 8 – Regressions on Size - OP sort

3-Factor Model

Sort on Size - OP (5x5)

Market ß	Low	2	3	4	High
Small	0.98	1.02	1.05	0.95	0.94
2	1.04	1.04	1.02	1.03	1.03
3	1.05	1.10	1.07	1.03	1.02
4	1.06	1.10	1.08	0.99	1.03
Big	1.11	1.10	0.98	0.98	0.94

t(ß)	Low	2	3	4	High
Small	25.777	28.245	31.193	21.714	18.680
2	31.318	31.715	26.416	32.961	26.280
3	31.019	27.012	27.902	29.498	23.355
4	24.243	28.025	24.705	26.079	26.167
Big	20.815	21.347	23.829	24.525	29.380

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.33	1.42	1.38	1.26	1.31
2	1.18	1.24	1.18	1.05	1.08
3	1.11	1.11	1.01	0.79	0.84
4	0.81	0.82	0.77	0.64	0.62
Big	0.41	0.30	0.34	0.16	-0.30

t(s)	Low	2.000	3.000	4.000	High
Small	15.251	18.709	18.244	14.192	12.656
2	16.339	16.338	12.940	14.828	12.648
3	13.406	12.193	10.738	9.901	10.270
4	8.729	7.692	7.782	7.352	7.037
Big	3.127	2.328	3.856	1.679	-4.534

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	0.04	0.08	0.02	0.15	0.22
2	0.02	0.10	0.05	0.00	0.21
3	0.07	-0.11	-0.11	-0.10	0.22
4	-0.08	-0.18	-0.16	0.01	-0.02
Big	0.19	0.10	0.05	0.04	0.16

t(h)	Low	2.000	3.000	4.000	High
Small	0.549	1.038	0.232	2.396	2.302
2	0.347	1.644	0.648	-0.052	3.257
3	0.804	-1.129	-1.649	-1.662	3.358
4	-1.170	-1.964	-1.682	0.126	-0.270
Big	1.585	0.740	0.710	0.498	3.270

R^2	Low	2.00	3.00	4.00	High
Small	0.94	0.94	0.94	0.92	0.88
2	0.94	0.95	0.94	0.94	0.93
3	0.93	0.92	0.93	0.93	0.91
4	0.91	0.92	0.90	0.91	0.90
Big	0.83	0.86	0.89	0.89	0.93

α	Low	2	3	4	High
Small	0.28%	0.13%	0.13%	0.09%	0.31%
2	-0.21%	-0.14%	-0.19%	-0.11%	-0.28%
3	-0.48%	-0.33%	-0.26%	-0.11%	-0.13%
4	-0.29%	-0.22%	-0.32%	-0.27%	-0.29%
Big	-0.03%	-0.10%	-0.22%	0.10%	0.27%

t(α)	Low	2	3	4	High
Small	1.088	0.505	0.501	0.346	0.873
2	-0.866	-0.614	-0.776	-0.474	-1.129
3	-1.793	-1.167	-1.018	-0.456	-0.470
4	-0.979	-0.784	-1.092	-1.065	-0.994
Big	-0.075	-0.271	-0.810	0.394	1.330

4-Factor Model

Sort on Size - OP (5x5)

Market ß

	Low	2	3	4	High
Small	0.99	1.01	1.04	0.95	0.94
2	1.04	1.04	1.02	1.03	1.03
3	1.05	1.09	1.07	1.03	1.02
4	1.06	1.09	1.07	0.99	1.03
Big	1.11	1.10	0.97	0.97	0.94

t(ß)

	Low	2	3	4	High
Small	25.823	28.474	30.584	21.698	18.752
2	30.685	31.904	26.541	32.110	26.454
3	30.286	26.542	27.162	28.655	23.534
4	23.807	28.155	24.734	25.539	26.204
Big	20.469	21.498	24.422	23.990	30.396

Factor loading SMB (s)

	Low	2.00	3.00	4.00	High
Small	1.32	1.45	1.40	1.26	1.30
2	1.20	1.27	1.20	1.07	1.08
3	1.14	1.14	1.03	0.81	0.83
4	0.82	0.86	0.81	0.66	0.62
Big	0.41	0.34	0.38	0.19	-0.32

t(s)

	Low	2.000	3.000	4.000	High
Small	14.911	19.427	19.008	14.437	12.176
2	17.043	18.570	12.942	16.447	12.913
3	14.328	12.958	10.950	10.111	10.415
4	9.168	8.295	8.161	7.521	6.687
Big	2.928	2.560	4.357	2.160	-5.091

Factor loading HML (h)

	Low	2.00	3.00	4.00	High
Small	0.02	0.16	0.08	0.16	0.18
2	0.07	0.18	0.10	0.04	0.21
3	0.15	-0.03	-0.05	-0.05	0.20
4	-0.03	-0.07	-0.06	0.07	-0.02
Big	0.20	0.18	0.16	0.13	0.11

t(h)

	Low	2.000	3.000	4.000	High
Small	0.200	1.767	0.961	2.145	1.587
2	0.910	3.095	1.052	0.745	2.901
3	1.727	-0.291	-0.638	-0.675	2.697
4	-0.380	-0.664	-0.560	0.725	-0.171
Big	1.266	1.176	1.894	1.614	1.900

Factor loading WML (w)

	Low	2.00	3.00	4.00	High
Small	-0.05	0.17	0.15	0.01	-0.11
2	0.11	0.19	0.11	0.11	0.00
3	0.19	0.18	0.13	0.11	-0.04
4	0.11	0.25	0.23	0.14	0.01
Big	0.03	0.19	0.25	0.23	-0.13

t(w)

	Low	2.000	3.000	4.000	High
Small	-0.680	2.489	1.958	0.079	-1.160
2	1.531	3.018	1.605	1.465	0.003
3	2.069	2.385	1.311	1.579	-0.503
4	1.144	2.345	2.410	1.556	0.087
Big	0.184	1.678	3.246	2.788	-2.519

R^2

	Low	2.00	3.00	4.00	High
Small	0.94	0.94	0.94	0.92	0.88
2	0.95	0.95	0.94	0.94	0.93
3	0.93	0.92	0.94	0.93	0.91
4	0.91	0.92	0.91	0.91	0.90
Big	0.83	0.86	0.90	0.90	0.94

α

	Low	2	3	4	High
Small	0.24%	0.27%	0.25%	0.10%	0.22%
2	-0.12%	0.01%	-0.10%	-0.03%	-0.28%
3	-0.33%	-0.18%	-0.16%	-0.02%	-0.16%
4	-0.20%	-0.02%	-0.14%	-0.16%	-0.28%
Big	-0.01%	0.06%	-0.02%	0.29%	0.17%

t(α)

	Low	2.000	3.000	4.000	High
Small	0.928	1.062	1.003	0.350	0.641
2	-0.470	0.034	-0.389	-0.111	-1.074
3	-1.325	-0.644	-0.640	-0.080	-0.593
4	-0.676	-0.074	-0.465	-0.651	-1.005
Big	-0.017	0.167	-0.082	1.087	0.834

5-Factor Model

Sort on Size - OP (5x5)

Market ß

	Low	2	3	4	High
Small	0.97	0.98	1.00	0.92	0.91
2	1.00	0.97	0.97	0.99	1.03
3	1.00	0.99	1.00	1.00	1.01
4	0.98	1.01	1.02	0.96	1.01
Big	1.05	0.97	0.92	0.90	0.96

t(ß)

	Low	2	3	4	High
Small	18.358	22.828	26.157	16.680	14.912
2	22.471	23.582	21.014	24.993	20.506
3	23.792	22.913	20.277	21.923	18.670
4	18.355	21.234	19.512	19.654	20.539
Big	15.776	18.400	17.733	18.631	22.472

Factor loading SMB (s)

	Low	2.00	3.00	4.00	High
Small	1.16	1.28	1.22	1.22	1.23
2	0.96	1.10	1.06	1.01	1.16
3	0.82	0.75	0.88	0.79	0.94
4	0.55	0.55	0.57	0.57	0.57
Big	0.02	-0.37	0.07	-0.05	-0.17

t(s)

	Low	2.000	3.000	4.000	High
Small	12.218	12.507	13.204	10.550	11.948
2	12.741	12.934	9.505	11.391	10.124
3	9.627	8.333	8.831	8.386	9.128
4	5.235	5.442	5.602	5.991	5.687
Big	0.157	-3.178	0.636	-0.535	-2.352

Factor loading HML (h)

	Low	2.00	3.00	4.00	High
Small	0.01	0.14	0.09	0.22	0.26
2	0.05	0.24	0.13	0.09	0.24
3	0.10	0.02	0.04	0.00	0.26
4	0.04	-0.06	-0.09	0.06	0.00
Big	0.20	0.19	0.09	0.15	0.16

t(h)

	Low	2.000	3.000	4.000	High
Small	0.098	1.458	1.088	2.183	2.045
2	0.605	3.507	1.407	1.208	2.587
3	1.129	0.213	0.436	0.020	2.843
4	0.525	-0.600	-0.874	0.644	0.046
Big	1.225	1.644	0.972	1.791	2.239

Factor loading RMW (r)

	Low	2.00	3.00	4.00	High
Small	-0.32	-0.38	-0.43	-0.17	-0.23
2	-0.52	-0.50	-0.38	-0.24	0.15
3	-0.66	-0.95	-0.48	-0.14	0.15
4	-0.73	-0.76	-0.53	-0.23	-0.16
Big	-0.84	-1.57	-0.63	-0.62	0.29

t(r)

	Low	2.000	3.000	4.000	High
Small	-2.548	-2.029	-2.777	-1.153	-1.094
2	-4.052	-3.470	-2.326	-1.332	0.895
3	-3.736	-6.197	-2.767	-0.941	0.978
4	-4.732	-4.075	-2.700	-1.255	-0.824
Big	-3.630	-9.357	-3.256	-3.580	2.803

Factor loading CMA (c)

	Low	2.00	3.00	4.00	High
Small	0.15	-0.18	-0.25	-0.25	-0.12
2	-0.07	-0.54	-0.30	-0.38	-0.13
3	-0.07	-0.44	-0.55	-0.39	-0.20
4	-0.41	-0.42	-0.24	-0.20	-0.09
Big	0.03	-0.23	-0.09	-0.43	-0.03

t(c)

	Low	2.000	3.000	4.000	High
Small	0.710	-1.016	-1.508	-0.970	-0.484
2	-0.417	-3.357	-1.592	-2.177	-0.632
3	-0.412	-2.507	-2.802	-2.203	-0.927
4	-1.824	-2.219	-1.178	-1.051	-0.415
Big	0.101	-1.087	-0.462	-2.087	-0.161

R^2

	Low	2.00	3.00	4.00	High
Small	0.94	0.94	0.95	0.92	0.88
2	0.95	0.95	0.94	0.94	0.93
3	0.94	0.94	0.94	0.93	0.91
4	0.92	0.93	0.91	0.91	0.90
Big	0.85	0.92	0.90	0.90	0.94

α

	Low	2	3	4	High
Small	0.31%	0.24%	0.27%	0.19%	0.38%
2	-0.08%	0.09%	-0.04%	0.02%	-0.28%
3	-0.33%	-0.04%	-0.04%	0.01%	-0.11%
4	-0.04%	0.03%	-0.16%	-0.18%	-0.23%
Big	0.13%	0.27%	-0.07%	0.33%	0.22%

t(α)

	Low	2	3	4	High
Small	1.345	0.961	1.113	0.697	1.098
2	-0.396	0.413	-0.187	0.099	-1.135
3	-1.329	-0.153	-0.159	0.036	-0.402
4	-0.154	0.123	-0.556	-0.735	-0.801
Big	0.345	1.119	-0.283	1.364	1.195

Appendix 9 – Regressions on Size – Inv sort

3-Factor Model

Sort on Size - Inv (5x5)

Market ß	Low	2	3	4	High
Small	0.97	0.97	1.05	1.00	1.03
2	1.02	1.06	1.01	1.05	1.03
3	1.07	1.02	1.05	1.05	1.07
4	1.04	1.02	1.05	1.05	1.06
Big	0.96	0.96	0.95	0.98	1.08

t(ß)	Low	2	3	4	High
Small	27.160	32.216	25.490	28.651	23.623
2	32.763	29.848	30.774	29.557	28.263
3	32.449	27.305	27.702	27.912	27.648
4	25.252	22.158	27.588	26.237	27.413
Big	14.597	22.237	24.448	25.776	31.931

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.34	1.37	1.42	1.38	1.15
2	1.17	1.15	1.18	1.17	1.11
3	0.96	1.06	0.94	0.95	1.03
4	0.70	0.84	0.63	0.72	0.72
Big	0.47	0.15	-0.05	-0.19	-0.07

t(s)	Low	2.000	3.000	4.000	High
Small	17.150	19.310	16.125	15.639	9.950
2	18.608	13.356	15.569	15.070	15.169
3	12.414	11.863	11.187	12.303	12.919
4	7.905	8.445	6.301	8.082	7.766
Big	3.164	1.932	-0.494	-2.405	-0.760

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	0.10	0.09	0.12	0.09	-0.12
2	0.13	0.01	0.11	0.10	-0.05
3	0.02	0.06	-0.01	-0.01	-0.07
4	0.06	0.03	-0.16	-0.09	-0.18
Big	0.41	0.28	0.37	0.00	-0.30

t(h)	Low	2.000	3.000	4.000	High
Small	1.661	1.319	1.486	1.243	-1.437
2	2.279	0.205	1.730	1.753	-0.910
3	0.294	0.861	-0.198	-0.180	-0.996
4	0.956	0.287	-1.904	-1.318	-2.275
Big	3.360	3.275	4.679	-0.010	-4.367

R^2	Low	2.00	3.00	4.00	High
Small	0.95	0.94	0.94	0.93	0.89
2	0.95	0.95	0.94	0.94	0.93
3	0.93	0.93	0.93	0.94	0.93
4	0.90	0.89	0.91	0.91	0.91
Big	0.81	0.87	0.90	0.92	0.92

α	Low	2	3	4	High
Small	0.23%	0.17%	0.30%	-0.14%	0.25%
2	-0.07%	-0.13%	-0.13%	-0.26%	-0.30%
3	-0.34%	-0.34%	-0.09%	-0.23%	-0.61%
4	0.14%	-0.26%	-0.21%	-0.32%	-0.42%
Big	0.25%	0.12%	0.32%	0.13%	-0.17%

t(α)	Low	2.000	3.000	4.000	High
Small	0.950	0.693	1.116	-0.511	0.739
2	-0.295	-0.556	-0.558	-1.040	-1.184
3	-1.414	-1.297	-0.370	-0.963	-2.270
4	0.488	-0.858	-0.762	-1.154	-1.518
Big	0.647	0.373	1.225	0.556	-0.671

4-Factor Model

Sort on Size - Inv (5x5)

Market ß

	Low	2	3	4	High
Small	0.98	0.96	1.05	1.00	1.03
2	1.01	1.06	1.00	1.05	1.02
3	1.07	1.02	1.05	1.05	1.06
4	1.04	1.02	1.04	1.05	1.06
Big	0.96	0.95	0.95	0.98	1.08

t(ß)

	Low	2	3	4	High
Small	27.244	31.841	25.343	28.265	23.721
2	32.652	29.884	30.607	29.246	28.032
3	31.716	26.807	27.321	27.775	27.025
4	25.145	22.292	27.097	26.058	26.949
Big	14.650	22.246	23.993	25.881	31.168

Factor loading SMB (s)

	Low	2.00	3.00	4.00	High
Small	1.33	1.40	1.43	1.38	1.14
2	1.19	1.17	1.21	1.18	1.11
3	0.98	1.09	0.96	0.95	1.05
4	0.73	0.87	0.67	0.72	0.74
Big	0.47	0.18	-0.05	-0.20	-0.06

t(s)

	Low	2.000	3.000	4.000	High
Small	16.715	20.293	16.209	14.869	10.041
2	19.623	13.971	17.245	15.996	15.852
3	12.797	12.672	11.545	12.627	13.542
4	8.315	8.812	6.680	7.740	7.994
Big	3.151	2.272	-0.465	-2.432	-0.586

Factor loading HML (h)

	Low	2.00	3.00	4.00	High
Small	0.09	0.15	0.15	0.09	-0.14
2	0.17	0.07	0.19	0.14	-0.03
3	0.08	0.14	0.04	0.00	-0.02
4	0.12	0.12	-0.07	-0.08	-0.12
Big	0.40	0.36	0.38	-0.02	-0.27

t(h)

	Low	2.000	3.000	4.000	High
Small	1.128	2.145	1.547	0.944	-1.421
2	2.652	0.958	2.917	2.288	-0.515
3	0.944	1.712	0.521	-0.042	-0.216
4	1.524	1.016	-0.666	-0.840	-1.225
Big	3.021	4.302	3.968	-0.243	-2.783

Factor loading WML (w)

	Low	2.00	3.00	4.00	High
Small	-0.04	0.14	0.07	-0.01	-0.04
2	0.09	0.13	0.18	0.10	0.04
3	0.14	0.18	0.12	0.02	0.12
4	0.14	0.21	0.21	0.03	0.14
Big	-0.03	0.20	0.01	-0.05	0.08

t(w)

	Low	2.000	3.000	4.000	High
Small	-0.585	2.202	0.825	-0.132	-0.460
2	1.490	1.920	2.760	1.680	0.478
3	1.839	2.137	1.527	0.214	1.589
4	1.730	1.995	2.070	0.295	1.377
Big	-0.240	2.391	0.153	-0.559	0.823

R^2

	Low	2.00	3.00	4.00	High
Small	0.94	0.94	0.94	0.93	0.89
2	0.95	0.95	0.94	0.94	0.93
3	0.94	0.93	0.93	0.94	0.93
4	0.90	0.89	0.91	0.91	0.91
Big	0.81	0.87	0.90	0.92	0.92

α

	Low	2	3	4	High
Small	0.20%	0.28%	0.35%	-0.15%	0.22%
2	0.01%	-0.03%	0.01%	-0.18%	-0.27%
3	-0.23%	-0.19%	0.00%	-0.22%	-0.51%
4	0.26%	-0.09%	-0.04%	-0.30%	-0.31%
Big	0.23%	0.28%	0.33%	0.09%	-0.10%

t(α)

	Low	2.000	3.000	4.000	High
Small	0.804	1.230	1.299	-0.555	0.639
2	0.032	-0.110	0.056	-0.707	-1.084
3	-0.944	-0.762	0.007	-0.907	-1.922
4	0.899	-0.290	-0.150	-1.071	-1.133
Big	0.595	0.880	1.275	0.377	-0.404

5-Factor Model

Sort on Size - Inv (5x5)

Market ß	Low	2	3	4	High
Small	0.98	0.94	1.01	0.94	0.97
2	0.99	1.02	0.96	1.00	0.93
3	1.04	0.97	1.01	1.00	0.97
4	1.04	0.98	1.02	1.00	0.98
Big	0.97	0.93	0.88	1.00	1.00

t(ß)	Low	2	3	4	High
Small	19.773	28.086	19.472	20.876	16.484
2	24.254	22.341	23.725	21.387	22.081
3	25.903	20.051	21.407	21.111	21.540
4	19.300	17.373	21.313	19.837	20.665
Big	10.786	16.334	19.555	18.938	26.045

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.22	1.20	1.31	1.25	1.10
2	1.00	1.03	1.07	1.11	1.03
3	0.69	0.84	0.87	0.92	0.91
4	0.51	0.63	0.43	0.60	0.61
Big	0.12	-0.11	-0.23	-0.14	-0.14

t(s)	Low	2.000	3.000	4.000	High
Small	13.808	11.975	12.036	13.516	7.795
2	14.058	10.494	12.205	10.429	13.577
3	8.154	9.025	9.120	10.298	11.603
4	4.416	6.888	4.040	6.044	6.209
Big	0.870	-1.040	-2.150	-1.551	-1.740

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	0.05	0.11	0.19	0.20	0.03
2	0.12	0.08	0.21	0.22	0.18
3	0.00	0.12	0.07	0.13	0.13
4	0.00	0.05	-0.16	0.00	0.01
Big	0.24	0.25	0.50	-0.03	-0.13

t(h)	Low	2.000	3.000	4.000	High
Small	0.568	1.393	1.726	2.073	0.286
2	1.529	0.887	2.613	2.939	2.969
3	0.004	1.588	0.893	1.665	1.737
4	0.010	0.494	-1.815	-0.012	0.062
Big	1.471	2.229	5.010	-0.277	-1.588

Factor loading RMW (r)	Low	2.00	3.00	4.00	High
Small	-0.18	-0.40	-0.33	-0.44	-0.33
2	-0.36	-0.35	-0.37	-0.31	-0.49
3	-0.54	-0.57	-0.29	-0.27	-0.57
4	-0.32	-0.48	-0.43	-0.39	-0.50
Big	-0.49	-0.52	-0.58	0.14	-0.41

t(r)	Low	2.000	3.000	4.000	High
Small	-1.462	-2.564	-2.063	-2.979	-1.586
2	-2.553	-2.284	-2.325	-1.876	-3.613
3	-3.543	-3.666	-1.984	-1.839	-3.898
4	-1.665	-2.270	-2.301	-2.126	-2.784
Big	-2.153	-2.549	-3.964	1.066	-2.748

Factor loading CMA (c)	Low	2.00	3.00	4.00	High
Small	0.24	-0.02	-0.23	-0.40	-0.63
2	0.10	-0.23	-0.37	-0.50	-0.90
3	0.15	-0.21	-0.32	-0.54	-0.81
4	0.29	-0.05	0.06	-0.35	-0.74
Big	0.78	0.18	-0.50	0.09	-0.68

t(c)	Low	2.000	3.000	4.000	High
Small	1.230	-0.165	-1.062	-2.066	-2.229
2	0.621	-1.257	-2.365	-2.694	-6.093
3	0.901	-1.114	-1.759	-2.743	-4.508
4	1.193	-0.227	0.332	-1.611	-3.858
Big	2.448	0.671	-2.434	0.423	-4.369

R^2	Low	2.00	3.00	4.00	High
Small	0.95	0.94	0.94	0.94	0.90
2	0.95	0.95	0.94	0.94	0.94
3	0.94	0.93	0.93	0.94	0.94
4	0.90	0.89	0.91	0.92	0.92
Big	0.83	0.88	0.91	0.92	0.93

α	Low	2	3	4	High
Small	0.21%	0.25%	0.42%	0.04%	0.47%
2	-0.02%	-0.01%	0.03%	-0.08%	0.00%
3	-0.27%	-0.17%	0.04%	-0.05%	-0.31%
4	0.14%	-0.15%	-0.14%	-0.16%	-0.15%
Big	0.17%	0.18%	0.55%	0.08%	0.08%

t(α)	Low	2	3	4	High
Small	0.971	1.062	1.605	0.163	1.318
2	-0.084	-0.043	0.131	-0.332	0.021
3	-1.208	-0.699	0.174	-0.212	-1.222
4	0.468	-0.503	-0.535	-0.594	-0.571
Big	0.497	0.538	2.441	0.376	0.333

Appendix 10 – Excess returns including Financial Companies

Excess Returns

Sort on Size - B/M (5x5)

Excess Returns

	Low	2	3	4	High
Small	2.55%	2.48%	2.38%	2.13%	2.22%
2	2.15%	2.15%	1.87%	1.63%	1.65%
3	2.18%	1.95%	1.60%	1.39%	1.27%
4	2.15%	1.45%	1.25%	1.26%	1.17%
Big	1.60%	1.20%	0.89%	0.57%	1.06%

Standard Deviation

	Low	2	3	4	High
Small	2.99%	3.56%	3.38%	3.46%	3.40%
2	3.56%	3.56%	3.25%	3.39%	3.46%
3	3.83%	3.19%	3.29%	3.28%	3.71%
4	3.54%	3.35%	3.62%	3.88%	4.11%
Big	3.09%	3.84%	3.60%	3.63%	3.36%

t-values

	Low	2	3	4	High
Small	2.829	2.312	2.329	2.042	2.162
2	2.006	2.006	1.905	1.594	1.584
3	1.883	2.028	1.610	1.403	1.137
4	2.018	1.431	1.146	1.075	0.942
Big	1.715	1.034	0.819	0.518	1.048

Appendix 11 – Summary Statistics including Financial Companies

Sort on Size - B/M (5x5)

Average of annual number of firms in Portfolio

	Low	2	3	4	High
Small	38	53	85	108	106
2	39	75	90	94	93
3	62	94	86	77	72
4	105	99	75	61	51
Big	146	71	55	51	69

Average of annual percentage of MV in portfolio

	Low	2	3	4	High
Small	0.26	0.37	0.6	0.72	0.71
2	0.52	0.89	1.03	1.05	1.08
3	1.23	1.78	1.57	1.35	1.29
4	3.63	3.27	2.38	1.95	1.66
Big	20.76	8.99	9.62	10.88	22.39

Average of Average Annual firm Size in Total MV

	Low	2	3	4	High
Small	2695	2904	2927	2728	2794
2	5686	4968	4647	4551	4739
3	8314	7739	7410	7166	7247
4	13980	13286	12782	12818	13348
Big	54573	47961	66932	91187	132342

Average of annual B/M ratios for portfolio

	Low	2	3	4	High
Small	0.05	0.13	0.21	0.34	0.59
2	0.06	0.13	0.21	0.33	0.59
3	0.06	0.13	0.21	0.34	0.60
4	0.06	0.12	0.21	0.33	0.62
Big	0.05	0.12	0.21	0.33	0.62

Average of annual E/P ratio in % for portfolios

	Low	2	3	4	High
Small	2.85%	1.50%	1.74%	1.94%	2.60%
2	2.05%	1.98%	2.24%	2.56%	3.14%
3	2.26%	2.33%	2.80%	3.18%	3.78%
4	2.42%	2.82%	3.32%	3.86%	4.83%
Big	2.91%	3.91%	4.52%	5.21%	6.93%

Average of annual D/P ratios in % for portfolio

	Low	2	3	4	High
Small	0.04%	0.17%	0.26%	0.34%	0.61%
2	0.25%	0.40%	0.51%	0.57%	0.68%
3	0.46%	0.53%	0.62%	0.70%	0.97%
4	0.51%	0.69%	0.90%	0.97%	1.44%
Big	0.63%	1.02%	1.20%	1.40%	2.05%

Appendix 12 – Regression on Size-BM sorts including Financial Companies

Two Factor Regression

Sort on Size - B/M (5x5)

Factor loading SMB (s)	Low	2	3	4	High
Small	1.14	1.42	1.33	1.45	1.41
2	1.31	1.19	1.21	1.26	1.26
3	1.04	1.07	1.10	1.11	1.16
4	0.87	0.92	0.80	0.83	0.77
Big	0.16	0.12	0.04	0.18	-0.03

t(s)	Low	2	3	4	High
Small	5.748	7.078	6.249	6.753	6.672
2	6.258	5.711	5.850	5.936	5.566
3	4.786	5.280	5.170	4.745	5.232
4	4.245	4.408	3.690	3.413	3.596
Big	0.741	0.524	0.157	0.871	-0.144

Factor loading HML (h)	Low	2	3	4	High
Small	-0.19	-0.08	0.12	0.28	0.40
2	-0.07	-0.22	-0.02	0.22	0.59
3	-0.27	-0.26	-0.02	0.23	0.61
4	-0.51	-0.21	-0.06	0.24	0.67
Big	-0.58	-0.21	0.17	0.45	0.84

t(h)	Low	2	3	4	High
Small	-1.033	-0.313	0.539	1.207	1.794
2	-0.349	-0.934	-0.109	0.990	2.527
3	-1.028	-1.080	-0.089	0.942	2.609
4	-2.129	-0.894	-0.278	0.899	2.977
Big	-2.586	-0.924	0.608	1.999	4.269

R^2	Low	2	3	4	High
Small	0.33	0.40	0.35	0.38	0.36
2	0.35	0.32	0.31	0.30	0.29
3	0.27	0.29	0.26	0.24	0.25
4	0.28	0.23	0.15	0.13	0.13
Big	0.08	0.00	-0.01	0.02	0.15

α	Low	2	3	4	High
Small	1.52%	1.29%	1.36%	1.10%	1.29%
2	1.05%	1.06%	0.87%	0.72%	0.95%
3	1.18%	0.94%	0.69%	0.61%	0.65%
4	1.17%	0.58%	0.56%	0.71%	0.90%
Big	1.16%	0.99%	0.95%	0.66%	1.54%

t(α)	Low	2	3	4	High
Small	1.879	1.519	1.601	1.314	1.554
2	1.242	1.224	1.035	0.845	1.081
3	1.356	1.109	0.797	0.688	0.734
4	1.378	0.685	0.637	0.775	1.000
Big	1.369	1.104	1.052	0.747	1.905

Three Factor Model Regrssion

Sort on Size - B/M (5x5)

Market ß	Low	2	3	4	High
Small	0.86	0.97	0.99	0.99	0.98
2	0.96	1.04	1.01	1.02	1.04
3	1.02	1.01	1.03	1.05	1.06
4	0.97	1.00	1.03	1.08	1.06
Big	0.99	1.06	1.05	1.04	0.93

t(ß)	Low	2	3	4	High
Small	18.01	24.64	33.61	32.66	35.74
2	25.30	27.51	39.70	36.70	40.56
3	22.11	30.87	31.43	32.21	34.69
4	31.46	24.26	27.43	27.18	30.65
Big	34.57	36.58	23.97	25.48	32.94

Factor loading SMB (s)	Low	2.00	3.00	4.00	High
Small	1.00	1.25	1.16	1.28	1.24
2	1.15	1.02	1.04	1.09	1.09
3	0.87	0.90	0.93	0.94	0.98
4	0.71	0.75	0.63	0.65	0.59
Big	-0.01	-0.06	-0.14	0.01	-0.19

t(s)	Low	2.00	3.00	4.00	High
Small	11.81	15.86	20.56	20.80	22.46
2	14.41	12.60	18.40	19.74	20.29
3	10.06	16.16	13.63	14.27	13.90
4	10.79	8.89	8.08	6.65	7.51
Big	-0.14	-0.84	-1.28	0.06	-2.69

Factor loading HML (h)	Low	2.00	3.00	4.00	High
Small	-0.25	-0.15	0.04	0.20	0.32
2	-0.15	-0.30	-0.10	0.14	0.51
3	-0.34	-0.34	-0.10	0.15	0.53
4	-0.58	-0.29	-0.14	0.15	0.58
Big	-0.66	-0.29	0.09	0.37	0.77

t(h)	Low	2.00	3.00	4.00	High
Small	-2.63	-1.95	0.91	3.50	7.28
2	-2.28	-4.73	-1.80	3.15	9.57
3	-3.42	-5.73	-1.91	2.12	9.54
4	-8.29	-3.89	-1.87	1.48	8.59
Big	-13.33	-4.86	0.85	4.89	13.30

R^2	Low	2.00	3.00	4.00	High
Small	0.83	0.94	0.96	0.96	0.96
2	0.90	0.95	0.96	0.96	0.97
3	0.92	0.95	0.94	0.95	0.95
4	0.92	0.92	0.91	0.92	0.93
Big	0.94	0.94	0.91	0.90	0.94

α	Low	2	3	4	High
Small	0.77%	0.44%	0.49%	0.24%	0.43%
2	0.21%	0.16%	-0.02%	-0.17%	0.04%
3	0.29%	0.06%	-0.21%	-0.31%	-0.27%
4	0.32%	-0.29%	-0.34%	-0.23%	-0.03%
Big	0.29%	0.06%	0.03%	-0.24%	0.72%

t(α)	Low	2	3	4	High
Small	2.109	1.778	2.331	1.201	2.220
2	0.654	0.720	-0.085	-0.844	0.231
3	1.063	0.277	-0.975	-1.468	-1.332
4	1.252	-1.105	-1.286	-0.880	-0.105
Big	1.478	0.292	0.113	-0.880	3.349

Sort on Size - B/M (5x5)

4-Factor Model

Market ß

	Low	2	3	4	High
Small	0.86	0.97	0.99	0.99	0.98
2	0.97	1.04	1.01	1.01	1.04
3	1.02	1.01	1.02	1.05	1.06
4	0.97	1.00	1.03	1.08	1.06
Big	0.99	1.06	1.05	1.04	0.93

t(ß)

	Low	2	3	4	High
Small	18.218	24.730	33.789	32.489	35.511
2	25.515	26.886	38.370	35.419	40.627
3	21.826	30.335	30.492	31.714	33.938
4	30.104	23.528	26.808	26.803	30.758
Big	32.910	37.073	25.710	25.349	31.381

Factor loading SMB (s)

	Low	2	3	4	High
Small	0.95	1.25	1.16	1.28	1.24
2	1.17	1.03	1.06	1.10	1.08
3	0.88	0.90	0.94	0.94	0.99
4	0.74	0.77	0.64	0.65	0.58
Big	0.00	-0.07	-0.18	0.01	-0.18

t(s)

	Low	2	3	4	High
Small	12.293	15.666	20.979	21.124	23.224
2	15.485	12.953	19.620	20.425	21.313
3	10.189	17.217	14.145	14.323	14.529
4	10.471	9.532	8.286	6.497	7.278
Big	0.083	-0.875	-1.821	0.132	-2.360

Factor loading HML (h)

	Low	2	3	4	High
Small	-0.40	-0.16	0.03	0.20	0.33
2	-0.16	-0.26	-0.06	0.17	0.50
3	-0.30	-0.31	-0.09	0.17	0.56
4	-0.54	-0.22	-0.10	0.18	0.57
Big	-0.61	-0.31	-0.02	0.39	0.77

t(h)

	Low	2	3	4	High
Small	-3.362	-1.785	0.499	3.344	7.423
2	-2.255	-3.815	-1.059	3.317	8.320
3	-2.785	-4.415	-1.368	2.180	9.729
4	-6.043	-2.585	-1.068	1.429	6.831
Big	-12.178	-3.878	-0.201	4.081	9.495

Factor loading WML (w)

	Low	2	3	4	High
Small	-0.36	-0.04	-0.04	0.00	0.02
2	0.01	0.10	0.10	0.05	-0.04
3	0.11	0.06	0.01	0.04	0.07
4	0.10	0.16	0.08	0.05	-0.04
Big	0.08	-0.05	-0.28	0.05	0.01

t(w)

	Low	2	3	4	High
Small	-3.538	-0.410	-0.611	-0.018	0.345
2	0.098	1.492	1.622	1.029	-0.779
3	1.287	0.806	0.192	0.640	1.192
4	1.040	1.797	1.018	0.622	-0.582
Big	1.473	-0.644	-2.685	0.507	0.119

R^2

	Low	2	3	4	High
Small	0.85	0.94	0.95	0.96	0.96
2	0.91	0.95	0.96	0.96	0.97
3	0.92	0.95	0.94	0.95	0.95
4	0.92	0.92	0.91	0.92	0.93
Big	0.94	0.94	0.92	0.90	0.93

α

	Low	2	3	4	High
Small	0.43%	0.41%	0.46%	0.23%	0.44%
2	0.07%	0.25%	0.08%	-0.10%	0.00%
3	0.39%	0.11%	-0.24%	-0.27%	-0.20%
4	0.44%	-0.15%	-0.26%	-0.18%	-0.07%
Big	0.32%	0.01%	-0.23%	-0.19%	0.68%

t(α)

	Low	2	3	4	High
Small	1.233	1.617	2.235	1.160	2.242
2	0.205	1.025	0.373	-0.496	0.021
3	1.356	0.495	-0.964	-1.206	-0.969
4	1.615	-0.549	-0.969	-0.702	-0.277
Big	1.495	0.066	-0.965	-0.684	3.245

Appendix 13 – Factor premia summary statistics for 1997-2017

Factor Returns													
Portfolio returns by Decile													
	Low/ Losser									High/ Winner			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**1-10**	**10-1**	**Average**
Size	2.00%	1.70%	1.31%	1.30%	1.15%	0.87%	1.04%	0.90%	0.85%	0.89%	1.11%	-1.11%	1.20%
B/M	1.63%	0.90%	1.00%	1.05%	0.97%	0.70%	0.76%	1.05%	0.77%	0.98%	0.65%	-0.65%	0.98%
Momentu	0.82%	1.01%	1.09%	0.94%	1.01%	1.06%	0.92%	0.99%	0.61%	0.91%	-0.10%	0.10%	0.94%
Inv	1.19%	1.19%	1.05%	1.20%	0.92%	1.18%	1.05%	0.91%	0.81%	0.74%	0.44%	-0.44%	1.02%
OP	1.20%	1.02%	1.02%	1.22%	0.91%	0.72%	1.03%	1.25%	0.92%	0.93%	0.28%	-0.28%	1.02%